



信息通信领域产学研合作特色期刊

第三届国家期刊奖百种重点期刊 | 中国科技核心期刊

ISSN 1009-6868

CN 34-1228/TN

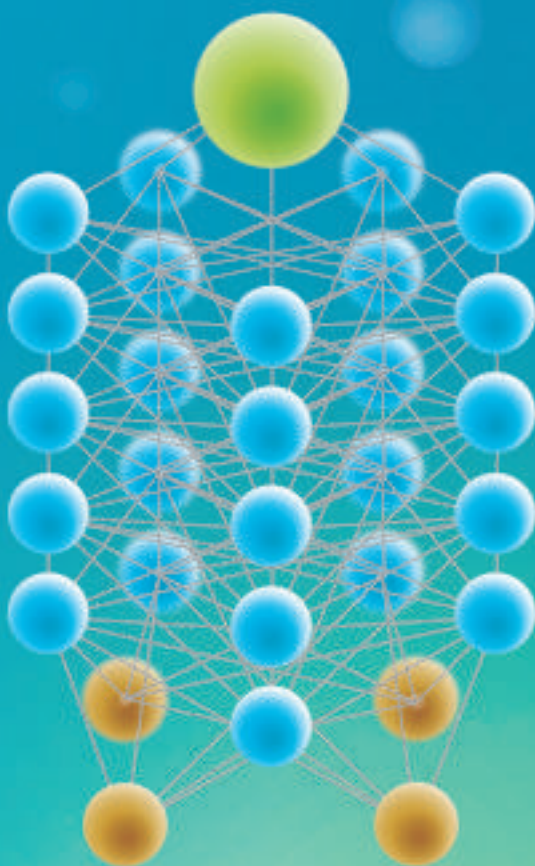
中兴通讯技术

ZTE TECHNOLOGY JOURNAL

tech.zte.com.cn

2017年8月 • 第4期

专题：神经网络与深度学习



Deep Learning



《中兴通讯技术》第7届编辑委员会委员名单

- 顾问

侯为贵（中兴通讯股份有限公司创始人） 钟义信（北京邮电大学教授） 陈锡生（南京邮电大学教授）
- 主任

陆建华（中国科学院院士,清华大学教授）
- 副主任

赵先明（中兴通讯股份有限公司总裁） 糜正琨（南京邮电大学教授）
- 副主任

马建国（广东工业大学教授） 陈前斌（重庆邮电大学副校长）

编委（按姓名拼音排序）

- 曹淑敏

中国信息通信研究院前任院长
- 陈建平

上海交通大学教授
- 陈杰

中兴通讯股份有限公司高级副总裁
- 陈前斌

重庆邮电大学副校长
- 葛建华

西安电子科技大学通信工程学院副院长
- 管海兵

上海交通大学电子信息与电气工程学院副院长
- 洪波

中兴发展股份有限公司总裁
- 洪伟

东南大学信息科学与工程学院院长
- 纪越峰

北京邮电大学信息光子学与光通信研究院
执行院长
- 蒋林涛

中国信息通信研究院科技委主任
- 李尔平

浙江大学信息学部副主任
- 李红滨

北京大学教授
- 李建东

西安电子科技大学副校长
- 李军

清华大学教授
- 李乐民

中国工程院院士,电子科技大学教授
- 李融林

华南理工大学教授
- 李少谦

电子科技大学通信与信息工程学院院长
- 李涛

南京邮电大学计算机学院院长
- 李星

清华大学教授
- 刘建伟

北京航空航天大学教授
- 陆建华

中国科学院院士,清华大学教授
- 马建国

广东工业大学教授
- 孟洛明

北京邮电大学教授
- 糜正琨

南京邮电大学教授
- 庞胜清

中兴通讯股份有限公司高级副总裁
- 孙知信

南京邮电大学现代邮政学院院长
- 谈振辉

北京交通大学教授
- 唐雄燕

中国联通网络技术研究院首席专家
- 童晓渝

中电科软件信息服务有限公司副总经理
- 王京

清华大学教授
- 王文东

北京邮电大学软件学院副院长
- 王翔

中兴通讯股份有限公司副总裁
- 王喜瑜

中兴通讯股份有限公司副总裁
- 卫国

中国科学技术大学教授
- 吴春明

浙江大学教授
- 邬贺铨

中国工程院院士
- 徐安士

北京大学教授
- 续合元

中国信息通信研究院技术与标准研究所总工
- 徐慧俊

中兴通讯股份有限公司执行副总裁
- 薛一波

清华大学教授
- 杨义先

北京邮电大学教授
- 杨震

南京邮电大学校长
- 尤肖虎

东南大学教授
- 张宏科

北京交通大学教授
- 张平

北京邮电大学网络技术研究院执行院长
- 张云勇

中国联通研究院院长
- 赵慧玲

中国通信标准化协会网络与业务能力技术工作
委员会主席
- 赵先明

中兴通讯股份有限公司总裁
- 郑纬民

清华大学教授
- 钟章队

北京交通大学教授
- 周亮

南京邮电大学通信与信息工程学院副院长
- 朱近康

中国科学技术大学教授



信息通信领域产学研合作特色期刊
第三届国家期刊奖百种重点期刊
中国科技核心期刊
工信部优秀科技期刊
中国五大文献数据库收录期刊
ISSN 1009-6868
CN 34-1228/TN
1995年创刊

办刊宗旨

以人为本,荟萃通信技术领域精英;
迎接挑战,把握世界通信技术动态;
立即行动,求解通信发展疑难课题;
励精图治,促进民族信息产业崛起。

Contents 目次

中兴通讯技术 总第135期 第23卷 第4期 2017年8月

专题:神经网络与深度学习

- 02 机器学习在大视频运维中的应用 屠要峰,吉铎,文韬
09 基于3D CNN的大规模视频手势识别研究 苗启广,李宇楠,徐昕
14 基于深度学习的多目标跟踪算法研究 陆平,邓硕,李伟华
20 车辆属性识别及跨场景标注方案 董振江,高燕,吴文熙
25 领域自适应目标识别综述 唐宋,叶茂,李旭冬
32 深度神经网络的结构基础:自动编码器与限制玻尔兹曼机 康文斌,彭菁,唐乾元
36 深度学习进展及其在图像处理领域的应用 刘涵,贺霖,李军

专家论坛

- 41 深度学习的能与不能 于剑
44 当深度学习遇到大视频数据 曾文军,罗翀
47 关于人机对话系统的思考 王小捷

企业视界

- 51 移动通信 Small Cell 基站供电解决方案与趋势分析 胡先红,刘明明

技术广角

- 56 基于运行大数据学习的复杂装备故障诊断技术及其典型应用 刘达新,裘乐淼,王志平
60 BRAS SDN&NFV 的演进思路 王怀滨,王延松

期刊基本参数:CN 34-1228/TN*1995*b*16*64*zh*P*¥ 20.00*15000*13*2017-08

Contents 目次

ZTE TECHNOLOGY JOURNAL Vol. 23 No. 4 Aug. 2017

Special Topic: Neural Network and Deep Learning

- 02 Applications of Machine Learning in Big Video O&M TU Yaofeng, JI Feng, WEN Tao
- 09 Large-Scale Video-Based Gesture Recognition Using 3D CNN Model
..... MIAO Qiguang, LI Yunan, XU Xin
- 14 Multiple Object Tracking Algorithm Based on Deep Learning
..... LU Ping, DENG Shuo, LI Weihua
- 20 A Solution for Vehicle Attributes Recognition and Cross-Scene Annotation
..... DONG Zhenjiang, GAO Yan, WU Wenxi
- 25 Domain Adaptation Object Recognition TANG Song, YE Mao, LI Xudong
- 32 Architectures of Deep Neural Networks: Auto-Encoders and Restricted Boltzmann Machines ...
..... KANG Wenbin, PENG Jing, TANG Qianyan
- 36 Advances in Deep Learning and Its Applications in Image Processing
..... LIU Han, HE Lin, LI Jun

Expert Forum

- 41 What Deep Learning Can or Cannot YU Jian
- 44 When Deep Learning Meets Big Video Data ZENG Wenjun, LUO Chong
- 47 Thoughts on Human-Computer Dialogue System WANG Xiaojie

Enterprise View

- 51 Power Supply Solutions and Trends Analysis for Small Cell Mobile Communication Base
Station HU Xianhong, LIU Mingming

Technology Perspective

- 56 Fault Diagnosis Technology for Complex Equipment Based on the Learning of Big Operation
Data and Its Typical Applications LIU Daxin, QIU Lemiao, WANG Zhiping
- 60 Evolution of BRAS SDN&NFV WANG Huaibin, WANG Yansong

敬告读者

本刊享有所发表文章的版权,包括英文版、电子版、网络版和优先数字出版版权,所支付的稿酬已经包含上述各版本的费用。

未经本刊许可,不得以任何形式全文转载本刊内容;如部分引用本刊内容,须注明该内容出自本刊。

2017年第1—6期专题

1 NB-IoT 体系标准与应用

王喜瑜 中兴通讯股份有限公司 副总裁
孙知信 南京邮电大学 教授

2 通信网络重构新技术

赵慧玲 中国通信标准化协会网络与业务
能力技术工作委员会主席

3 5G 无线传输技术

卫国 中国科学技术大学 教授

4 神经网络与深度学习

焦李成 西安电子科技大学 教授
杨淑媛 西安电子科技大学 教授

5 硅基光电子集成技术与器件

陈建平 上海交通大学 教授

6 虚拟现实技术与应用

潘志庚 杭州师范大学 教授



焦李成

西安电子科技大学教授、博士生导师，智能感知与图像理解教育部重点实验室主任，智能感知与计算国际联合研究中心主任，智能感知与计算国际合作联合实验室主任，国家自然科学基金二等奖获得者，享受国务院政府津贴的专家，中国人工智能学会副理事长，教育部科技委学部委员，国务院学位委员会学科评议组成员，并担任国际权威期刊IEEE TGRS的副主编；长期从事智能感知与计算、图像理解与目标识别、深度学习与类脑计算等方面的研究；撰写了中国首部神经网络专著——《神经网络系统理论》和深度学习专著——《深度学习、识别与优化》，所发表的论著被他人引用超过25 000篇次，H指数为65。



杨淑媛

西安电子科技大学教授、博士生导师，智能感知与图像理解教育部重点实验室骨干，类脑智能与深度学习研究中心主任，教育部新世纪优秀人才，陕西省科技新星，吴文俊人工智能青年创新奖获得者；发表论文20余篇，授权国家发明专利21项，合作出版专著3部。

导读

作为联接主义智能实现的典范，神经网络采用广泛互联的结构与有效的学习机制来模拟人脑信息处理的过程，是人工智能发展中的重要方法，也是当前类脑智能研究中的有效工具。目前，模拟人脑复杂的层次化认知特点的深度学习已经成为类脑智能中的一个热点研究方向，无论是在学术界还是工业界都引起了极大的研究兴趣和投入，并且颠覆和革新了许多传统的技术和理念。通过增加网络的层数和节点数、采用全新的网络结构、设计高效的学习优化策略，所构造的“深层神经网络”能够使机器获得从海量数据中学习“抽象概念”和“复杂规律”的能力，在诸多领域都取得了巨大的成功，又一次掀起了神经网络研究的一个新高潮。

近年来，深度学习技术发展日新月异，各种网络模型和算法不断涌现，在越来越多的领域得到了成功的应用，并且较传统方法在性能上实现了飞跃。深度神经网络的功能已经从传统的分类识别扩展到了特征提取、分布学习、端到端的建模等方面，学习训练方法也从简单的反向传播发展出了预训练加微调、结构的动态调整、数据知识的迁移等策略，应用领域也从传统的视觉任务延伸至语音识别、机器翻译和诸如AlphaGo的特定智能问题。除了对神经网络自身技术的革新外，深度学习的出现一方面源于计算机存储和计算能力的发展，另一方面则依赖于海量数据的高效获取和标记技术。目前，包括FPGA、GPU乃至TPU在内的并行计算技术的发展，为深度学习的高效实现提供了良好的平台支撑，另一方面包括互联网、遥感、医学等在内的大数据则为深度学习提供了丰富的数据源。因此，无论是学术界还是工业界都具备了进行研究和应用的软硬件条件，这必将提升数据处理和理解的智能化水平，产生有影响力的学术成果和技术产品。

尽管深度学习已经取得了许多瞩目的成就，但是其仍然存在一些尚未解决的问题。目前在网络的结构设计、参数设置和学习优化等方面缺少理论指导，很多时候往往需要凭借人为经验和主观判断来设计出适用于特定问题的模型。根据领域知识，设计出适用于特定问题的网络模型和算法是当前发展的必然趋势。虽然深度学习能对问题实现端到端的建模，但随之而来的是复杂和繁琐的训练过程，训练过程中的技巧性和策略性很强，对最终网络的性能起着至关重要的作用。在有限标记数据情况下，借助于外部的海量数据来实现知识的迁移是解决小样本问题的有效途径。同时，关于深度学习潜在的“过拟合”问题，目前尚无系统性的研究和结论，模型的泛化能力值得进一步关注。从网络的约简和正则化等角度开展研究是今后深度学习一个很重要的研究方向，诸如稀疏性和选择性的引入将会使网络更加智能化。

本期专题我们将和大家一起讨论关于深度学习在学术界和工业界的一些最新进展，包含了来自高校和公司的研究人员在这方面所做的创新性研究工作。这些论文凝聚了作者多年的研究成果和工作经验，希望能给读者有益的启示与参考。在此，对各位作者的积极支持和辛勤工作表示衷心的感谢。

焦李成 杨淑媛

2017年6月25日

机器学习在大视频运维中的应用

Applications of Machine Learning in Big Video O&M

屠要峰/TU Yaofeng
吉锋/JI Feng
文韬/WEN Tao

(中兴通讯股份有限公司, 江苏 南京
210012)
(ZTE Corporation, Nanjing 210012, China)

随着移动互联网和宽带网络的快速发展, 视频业务以广泛的受众、高频次的使用、较高的付费意愿, 已经具备成为“杀手应用”的潜质。越来越多的电信运营商将视频业务视为发展的新机遇, 并作为与宽带、语音并列的基础业务。据 Conviva 用户视频报告的数据, 35% 的用户把视频观看体验作为选择视频服务的首要条件^[1]。因此, 运维保障成为视频业务的关键。

当前视频业务发展已进入“大内容”“大网络”“大数据”“大生态”的大视频时代。业务形态多样, 包括交互式网络电视 (IPTV)、基于互联网应用服务 (OTT) 的 TV、移动视频等; 组网复杂, 视频在多屏之间的无缝衔接、码率格式适配等需求对网络提出了更高的要求; 数据多样性大大增加, 需要从视频码流、终端播放器、内容分发网络 (CDN)、业务平台、网络设备各个环节获取数据, 既有结构化数据, 又有半结构化、非结构化数据; 数据实时性要求大大提高, 传统网管采集数据的粒度是 5 min, 而大

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0002-007

摘要: 通过对中兴通讯大视频运维系统整体架构和关键模块的介绍, 以及机器学习技术在大视频运维系统中端到端异常检测、根因分析与故障预测等场景的具体应用的分析, 并结合硬盘故障预测的实例, 认为随着人工智能在运维领域的应用发展, 从基于规则的自动化运维转向基于机器学习的智能运维必然成为趋势。中兴通讯适时采用了机器学习方法来提取历史巡检数据中蕴含的故障特征, 并构建集成预测模型来提升大视频运维的精度和效率, 目前取得了较好的效果。

关键词: 大视频; 大数据; 机器学习; 人工智能

Abstract: In this paper, the overall architecture and key modules of ZTE big video operation and maintenance (O&M) system are introduced. Then the application scenes of machine learning technology in this system, including the end-to-end anomaly detection, root cause analysis and fault prediction are analyzed, as well as the instances of hard disk breakdown prediction. With the development of artificial intelligence technologies in the field of O&M, the trend which is from rule-based automatic to machine learning-based intelligent will be formed. To improve the precision and efficiency of the big video O&M, ZTE has adopted a machine learning method to extract the fault features contained in the historical inspection data, and has built the integrated prediction model, which has achieved good performances.

Keywords: big video; big data; machine learning; artificial intelligence

视频业务要求秒级的数据采集和分析, 数据量和计算量增加了百倍。

这些都对传统的运维模式和技术方案带来很大的挑战。如何在大数据背景下客观评价和度量终端用户的体验质量, 如何界定视频业务系统故障和网络故障, 如何快速诊断网络中的故障并提前发现网络隐患, 如何发掘视频业务运营和利润的增长点, 成为各大运营商对大视频业务运维的关注重点。

1 大视频智能运维系统的架构及关键技术

在原有运维技术手段基础上, 进一步依托大数据及人工智能技术, 对

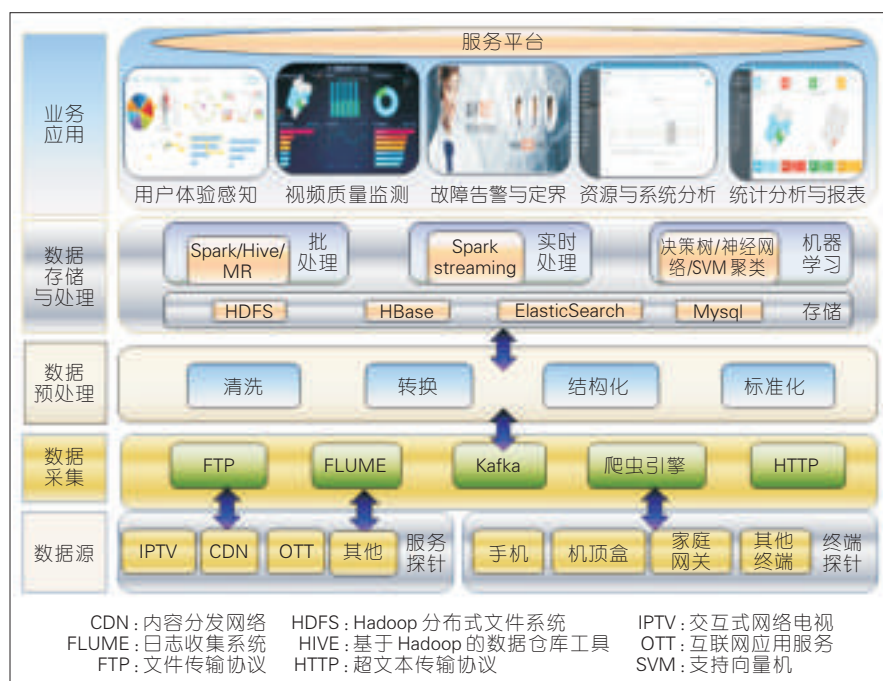
大视频业务系统产生的各类信息进行汇聚、分析、统计、预测等, 中兴通讯形成了智能化的大视频运维系统, 其系统架构如图 1 所示。

大视频运维系统有以下几个部分组成:

(1) 数据源。数据源主要指大视频业务智能运维所需要采集的数据, 包括终端的播放记录、关键绩效指标 (KPI) 数据; 接入网络的用户宽带信息、资源拓扑数据; CDN 的错误日志、告警、链路状态、码流信息等; IPTV 业务账户、频道/节目信息等。

(2) 数据采集及预处理。数据采集层主要是 Kafka、文件传输协议 (FTP)、超文本传输协议 (HTTP) 等用

收稿日期: 2017-05-28
网络出版日期: 2017-07-06



▲ 图1 大视频运维系统架构

于数据采集的组件;数据预处理是指对各种异构日志数据进行解析、转换、清洗、规约等操作,主要完成数据使用前的必要处理及数据质量保证。

(3)数据分析处理。数据分析处理主要包括流式计算处理框架Spark、离线批处理MR框架、人工智能计算框架、数据存储及检索引擎等。业务组件包括批处理、数据实时分析、机器学习等模块。批处理模块主要是对时效性要求不高的业务模块的处理及数据的离线分析,包括但不限于故障及异常的根源分析、故障及特定规则阈值的动态预测、事件的依赖分析及关联分析、异常及重要时序模式发现、多事件的自动分类等;数据实时处理主要是对于时效性要求较高的安全事件进行监测控制、异常检测与定位、可能引发严重故障的预警、对已知问题的实时智能决策等;机器学习模块包括离线的机器学习训练平台、算法框架和模型。

(4)业务应用层。业务应用层主要提供智能业务监测控制、端到端故障定界定位、用户体验感知、统计分析与报表等主要业务场景的分析及

应用。

大视频运维系统涉及的关键技术包括:

(1)大数据技术。该技术可以构建基于大数据的处理平台,实现数据的采集、汇聚、建模、分析与呈现。

(2)探针技术。该技术可以实现全网探针部署,包括机顶盒探针、直播源探针、CDN探针、无线探针、固网视频探针等,通过探针技术实现全面的视频质量实时监测控制以及数据采集。

(3)视频质量分析指标。该指标以用户体验为依据建立视频质量评估体系,对视频清晰度、流畅度、卡顿等多项用户体验质量(QoE)指标进行分析。

(4)人工智能技术。机器学习本身有很多成熟的算法和系统,以及大量的优秀的开源工具。如果成功地将机器学习应用到运维之中,还需要3个方面的支持:数据、标注的数据和应用^[9]。大视频系统本身具有海量的日志,包括从终端、网络、业务系统多方面的数据,在大数据系统中做优化存储;标注的数据是指日常运维工

作会产生标注的数据,比如定位一次现网事件后,运维工程师会记录下过程,这个过程会反馈到系统之中,反过来提升运维水平;应用指运维工程师是智能运维系统的用户,用户使用过程发现的问题可以对智能系统的优化起正向反馈作用。

2 人工智能技术在大视频运维系统中的应用

2.1 基于人工智能的端到端智能运维

传统电信网络、业务系统的运维模式通常是在故障发生后,运维、开发人员被动地进行人工故障的定位与修复。技术专家通过分析系统日志,依据事先制订的系统运行保障规则、策略和依赖模型,判断故障发生的原因并进行修复。这一过程不仅工作量巨大,操作繁琐,代价高昂,容易出错,且不能满足持续、快速变化的复杂系统环境需求。

大视频业务系统的故障定界定位尤其复杂且耗时耗力,原因在于:大视频系统中网元众多且业务流程复杂,如包括IPTV管理系统、电子节目菜单(EPG)、CDN、机顶盒、直播源编码器等众多网元,发现问题需要各个网元一起定位排查,对人员技能的要求很高。大视频系统对网络要求比较高,机顶盒经过光网络单元(ONU)、光线路终端(OLT)、宽带远程接入服务器(BRAS)、核心路由器(CR)等,从接入设备、承载设备到CDN服务器,中间任何一个网络设备出现丢包、抖动等问题都会导致用户的观看体验受影响,对这种卡顿分析是一个大难题。随着视频业务的快速发展和业务量不断增长,如何快速定位问题,降低运维门槛变得越来越迫切。

端到端智能运维系统就是利用大数据采集分析、人工智能与机器学习等技术提升系统运维智能化能力,从智能化的故障定位、智能化的根因分析机制入手,覆盖从被动式事后根

源追溯到主动式事中实时监测控制及事前提前预判的各种业务场景(如图2所示),提供从数据收集分析,故障预判到定位,再到故障自动修复的端到端保障能力。

面向历史的事后追溯主要有历史故障根因分析、系统瓶颈分析、业务热点分析等;面向实时的事中告警主要有异常监测、异常告警、事件关联关系挖掘、实时故障根因分析等;面向未来的事前预判主要有故障预测、容量预测、趋势预测、热点预测等。其中,事后追溯更多面向离线、非实时的运维故障分析,事中告警和事前预判更多面向实时或准实时的运维故障检测、分析及预测。

机器学习技术在端到端的智能运维系统中有几个应用点。

(1) 日志预处理模块

预处理的核心问题是将半结构、非结构化的日志转换为结构化的事件对象。事件被定义为一种现实世界系统状态的体现,通常涉及到系统状态的改变。本质上,事件是时序的且经常以日志的方式进行存储,例如:业务事务日志、股票交易日志、传感器日志、计算系统日志、HTTP请求、数据库查询和网络流量数据等。捕获这些事件体现了随着时间变化的系统状态和系统行为以及它们之间的时序关系。事件对象可以简单定义为: $Event = \{ \text{时间戳}, \text{事件类型}, \langle \text{属性1: 属性值1}, \text{属性2: 属性值2} \dots \rangle \}$ 。事件挖掘是一系列从历史事件和日志数据中自动、高效地获取有价值知识的技术,正确提取事件才能后续从时间、空间等多角度挖掘事件之间的关联、依赖等关系。将文本日志集转换成系统事件的典型技术方案包括:基于日志解析器、基于分类和基于聚类的方法。

最为直接的解决方案是采用日志解析器,该方法为每一种特定的系统日志实现对应的日志解析器,每种类型的日志采用正规表达式或预定义模板进行抽取。这种方式需要用

户了解系统日志,一个日志解析器难以适配不同格式的多种系统日志,需要大量的人力来开发定制的日志解析器软件。从机器学习辅助人工完成日志解析的角度,可以采用分类或聚类的方式。

日志分类方法是一种直接从日志数据中识别事件类型的方法,它通过分类器模型将一条条日志消息划分成若干个预定义的事件类型,如图3所示。

一种简单的分类方法是为每一个事件类型预先定义一种对应的正则表达式模式(如前所述的解析器或称为过滤器);另一种更为通用的日志消息分类方法是基于机器学习分类模型的方法,即用户提供一些标记过的日志消息,每个消息的事件类型已被明确标注;然后,机器学习算法根据标记的数据建立一个分类模型,利用这个模型对新的日志消息进行分类。虽然这种方式带来一定的泛化性,其主要问题在于需要大量的带标记日志消息,需要一定数据积累与人力消耗。

另外一种基于聚类的方法,不需要大量人力且适用于多种系统日志,虽并非十分准确,但可以应用于能够容忍一些错误或噪音事件的事件挖掘应用中。日志消息聚类采用无监

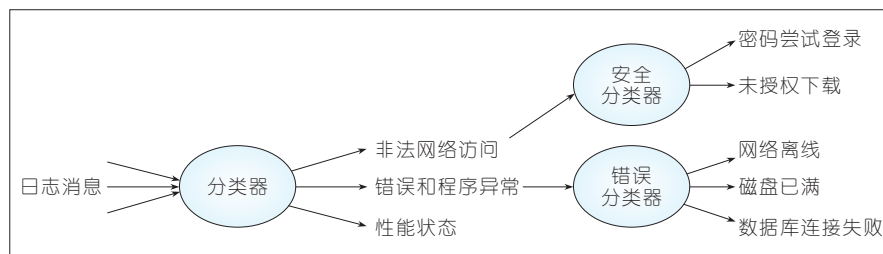
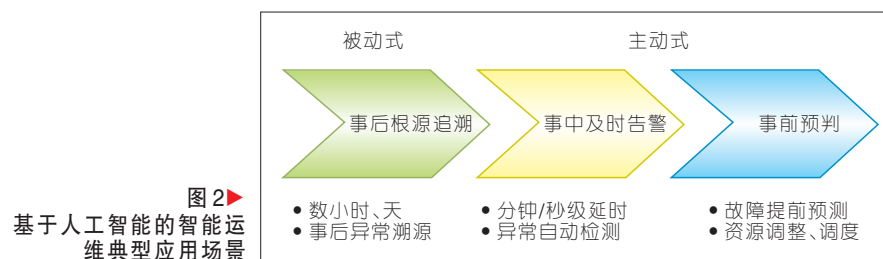
督方法将日志划分为各种类型事件,因为日志消息聚类并不要求准备一系列标记过的训练数据,所以这种方法更加实用。业界典型的日志聚类方式包括基于日志消息签名^[3]和基于树状结构的聚类^[4]等算法。

在实际应用场景中,需要根据业务的复杂度和数据积累情况,综合选择日志预处理解决方案:准确性要求非常高的场景需要使用专业日志解析器;而分类、聚类的机器学习方式更适合容忍一些错误或噪音事件的事件挖掘应用。

(2) 日志离线分析模块

日志离线分析的核心问题是通过机器学习算法发现事件之间的关联、依赖关系。离线分析负责从历史日志数据中获得事件间关联性和依赖性知识并构建知识库。事件挖掘综合利用数据挖掘、机器学习、人工智能等相关技术去发现事件之间的隐藏模式、未来的趋势等关系。分析人员可以利用已发现的事件模式对未来事件的行为作预测,同时挖掘出的事件依赖关系也可以用于系统故障的诊断,帮助运维人员找出问题的根因,达到解决问题的目的。

事件的关联分析,本质上根据日志文件中的每个消息事件的时间戳,发现时序事件之间的关联性。我们



▲图3 日志消息中事件类型分类示例

重点挖掘两种类型的相关性:基于时序连续值数据的相关性和基于离散事件数据的相关性。以大视频系统为例:中央处理器(CPU)使用率、内存使用情况、磁盘读写数据量、网络接收或发送数据量都可以表示为时间序列的连续值;而应用程序服务器上请求和应答序列被视作事件数据,因为每个数据项的值都是属于某个类别的离散值,例如:CPU使用率的时序图与磁盘读写数据量时序图(如横轴时间、纵轴数值的可视化表示)就会有很强的相关性;网络异常事件与应用程序服务器上请求和应答异常事件,通过离散事件数据分布图(横轴时间、纵轴事件类型的可视化展示)分析就存在一定相关性,如两类事件基本在同一时间点同时出现,具备一定的关联性。

事件之间存在关联性,不一定表明事件之间一定存在依赖或因果关系,比如事件A和B具备相关性,并不代表A引发B或B引发A,因此需要基于关联关系基础上进一步挖掘相互依赖关系。所谓事件依赖分析,发现类似 $A \rightarrow B$ 的依赖关系,最终形成一个事件依赖的动态概率模型图。如图4所示,A、B、C对应于大视频运维中不同事件,通过基于时间窗的事件依赖算法^[5]挖掘出各种故障事件之间的依赖并形成相应的依赖图。

总之,离线分析主要是通过机器学习算法形成关联、依赖的规则或概率图模型,另外还包括利用历史时序故障数据进行传统机器学习特征工

程建模或深度学习端到端的时序建模,为接下来的在线实时故障分析、定位与预测等提供支撑。

(3) 实时分析模块

实时分析模块负责实时处理新产生的日志数据并根据离线分析获得的知识模型完成在线运维的管理操作。典型的实时分析技术主要有异常检测、故障根因分析、故障预判和问题决策等。

对于实时的异常检测,可选择的方案有两种:基于监督学习和基于无监督的方案。前者利用基于离线训练出来的检测模型进行判断,这种方式不如后者使用普遍。

故障预判更多的是基于历史的数据进行分析建模,如下文即将讲述的基于Smart硬盘故障预测示例。

问题决策则更为全面,对前述检测出的异常,预判即将要出现的故障以及定位已知故障的原因,进行高层的资源调度,或发出设备替换的决策指令,最终避免可能出现的故障,自动修复已知的故障(若可以修复)或者发出告警通知运维人员进行人工修复。

(4) 智能故障定位及根源分析

故障智能定位是模拟人工排查故障的流程,对可疑的故障检查点进行逐一排查,通过采集各业务模块的告警、性能指标、错误和异常日志,组织生成故障定位的基础事件数据,针对故障现象配置对应的检查点及处理建议。

在故障定位时,从故障现象出

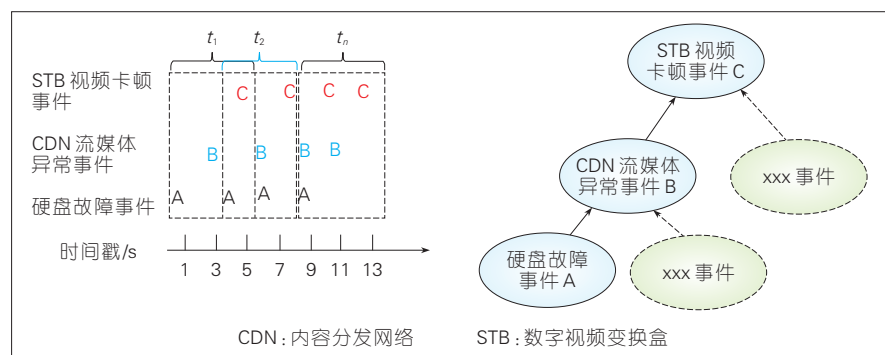
发,通过中序遍历方式遍历整个故障树。前一个节点的出参是后一个节点的入参,检查点调用应用程序编程接口(API)检查本节点的故障原因是否存在,通过API来分别从各种网元获取对应现象的证据信息,直至分析到叶子节点。然后将所有满足条件的节点进行回归,根据权重返回现象的原因。遍历结束后综合各个节点的检查结果形成本次故障定位的诊断结论。

在用户报障时可能对故障产生的时间、触发的位置、观看的节目等信息记忆模糊不清。在故障定位过程中首先需要从用户的行为记录里筛选出故障记录,这个筛选的过程采用前述的日志聚类算法,对影响用户感知的KPI进行聚类,对聚类结果根据预定的规则或分类器判断出属于故障类的记录。如果有多条故障记录,任选一条故障记录进行定位。除通过算法筛选故障记录外,另提供人工辅助筛选功能提高准确性。

智能化的根因分析,主要根据前期分析出来的事件依赖概率图模型,建立基于历史故障定位及处理经验集的专家知识库,利用机器学习的理论与技术,在多维变量间因果关系做出权重的动态调整,调整各个检查点的权重。故障定位方法较之前的传统方法具有更精确的错误定位效果和更显著的定位效率。

2.2 基于人工智能的硬盘故障预测实例

当前大视频运维过程中遇到的难题之一就是CDN故障硬盘的置换。为了规避软硬件风险,提升数据中心管理效率,制订合理的数据备份迁移计划,业界各大主流IT企业均展开针对硬盘故障预测的研究工作。研究者认为:在此预测技术的支撑下,可以极大地提升服务/存储系统的整体可用性。我们接下来将列举一个基于机器学习实现的CDN硬盘故障预判的实例。



▲ 图4 基于时间窗的依赖关系挖掘(左)与依赖概率(右)

当前,自我监测分析和报告技术(SMART)已经成为工业领域中硬盘驱动状态监测和故障预警技术的事实标准^[6]。研究表明:硬盘的一些属性值如温度、读取错误率等,和硬盘是否发生故障有一定的关系。如果被检测的属性值超过预先设定的一个阈值,则会发出警报。然而,硬盘制造商估计,这种基于阈值的算法只能取得3%~10%的故障预测准确率和低预警率^[7]。学术界和工业界在采用机器学习方法提升SMART硬盘故障预测精度方面的工作由来已久,但受限于数据集规模,现有方法取得的预测模型效果不佳。近年来,随着越来越多厂商的关注,基于SMART巡检数据的硬盘故障预测研究有了很好的数据支撑,一方面体现在硬盘规模快速增长,另一方面体现在采样工作正规化。在以上高质量数据支撑下,基于SMART巡检数据的故障预测水平得到了显著提升。

我们在大视频运维中基于SMART数据进行硬盘故障预测,采用了基于旋转森林的集成预测模型方案,基本流程如图5所示。

将SMART扫描数据集按照局点和硬盘型号进行细分,每个局点每个

硬盘型号的数据分别建立预测模型,每个预测模型的构建过程为:

(1)特征工程。特征工程是决定预测效果的关键步骤。我们不但需要考虑观测点当时的SMART取值,也需要考虑该SMART取值的历史变化趋势、震荡幅度、跳变频率等因素,主要策略包括取高价值属性和衍生时序特征。取高价值属性,即采用“数据驱动和领域知识相结合”的策略,一方面和相关硬件专家交流,听取他们的领域指导意见;另一方面,从故障硬盘的历史SMART记录集出发,找出“故障硬盘和健康硬盘在该属性上统计性质存在不一致”的SMART属性。专家知识和数据驱动结果都作为特征工程结论的一部分,宁多勿少。衍生时序特征,即在找出具有提示性效果的高价值SMART属性后,对其时序特征做进一步衍生、调整。以上两种特征工程策略相互补充,共同组成了模型训练需要的特征空间。

(2)模型训练。模型选择和训练、优化是构造预测模型的直接步骤,由于基于SMART记录集做硬盘预测是一个高维分类问题,同时正负数据严重不平衡,采用线性分类模型

往往没有很好结果,因此考虑采用构造非线性模型来解决问题,主要分两大步骤:重新平衡正负样本和非线性建模。重新平衡正负样本,即采用“过采样+降采样”结合的策略,对于负样本(健康硬盘),考虑采用聚类方法提取聚类质心,将质心附近的样本按比例提取作为该聚类的代表,从而实现降采样,而聚类算法和聚类质量评价准则需要根据实际数据分布来决定;对于正样本(故障硬盘),考虑采用过采样方法来提升正样本数据分布。以上降采样和过采样策略结合,把正负样本的比率从1:50重构到1:5以内,重构训练集。非线性建模,即利用旋转森林技术对以上训练集进行降维,并选择核方法、神经网络来构造分类超平面,择优选择其中有代表性的模型,然后再将这些模型利用层叠泛化技术组合形成最后的预测模型。

(3)模型评估。模型评估即将模型训练阶段生成的模型,在测试集上进行测试,重点关注预测准确率和故障覆盖率(召回率),直到选出符合要求的模型。

(4)模型上线。模型上线即将通过模型评估的最终模型部署到现网

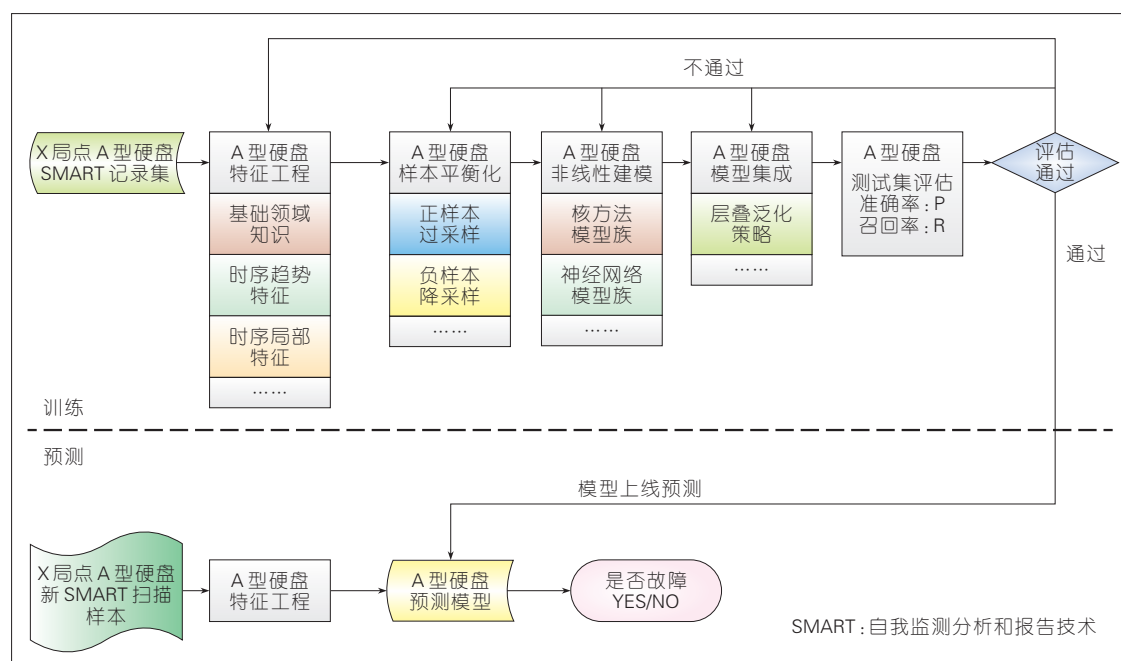


图5 硬盘故障预测流程

环境后,每次 SMART 扫描得到的新样本均需要输入本模型,得到“未来一段时间该硬盘是否发生故障”的预测结果。

我们在 Backblaze 数据集(2016 年 Q1—Q2)^[8]上选取了某型号希捷硬盘的 SMART 扫描记录做验证,其中时间跨度为 6 个月 182 天,数据粒度为每天扫描,涉及到 11 890 块硬盘(连续 3 天扫描找不到则视为故障盘,这种盘共 242 块),共 2 118 925 条扫描记录。按照 7:3 的大致比例划分训练集和测试集:测试集共 3 560 块硬盘(故障盘共 83 块),共 634 950 条扫描记录。

我们采用如表 1 所示的 SMART 基础属性来进行建模,共有 12 个基础属性。

考虑到 SMART 属性前后变化趋势也可能昭示着后续硬盘故障,因此我们在以上基础属性上衍生了时序率属性,包括每个基础采集之前 9 天内的相对变化率。

将以上基本 SMART 属性和衍生属性融合,作为 SMART 故障预测的特征参与模型构建。我们采用的子分类器如表 2 所示,共 4 类 18 个。

随着旋转森林特征子集分块参数的变化,生成的子分类器对故障的

▼表 2 旋转森林备选异构子分类器

类型编号	学习器	备注
①	多层感知机 MLP	单隐层节点数目分别为 6/7/8/9/10/11/12/13/14
②	支持向量机 SVM	高斯核函数, Cost =10, σ 分别为 0.6/0.7/0.8
③	支持向量机 SVM	高斯核函数, Cost =100, σ 分别为 0.6/0.7/0.8
④	支持向量机 SVM	高斯核函数, Cost =1 000, σ 分别为 0.6/0.7/0.8

MLP:多层感知机

SVM:支持向量机

预测能力也在不断调整,最终生成不同分块参数对应的模型在同一测试集下不同的预测效果(如图 6 所示)。当旋转森林特征子集分块参数为 6 时,能够取得 98.80% 的最高覆盖率,同时达到 5.75% 的误报率;当旋转森林特征子集分块参数为 5 时,能够取得 3.57% 的最低误报率,同时达到 97.6% 的覆盖率。

此外,在当前测试中可以发现:绝大部分故障在预报 30 天内可以被证实,图 7 是预警提前天数的分布累计情况。

综上所述,基于 SMART 的故障预测技术在当前智能运维领域已经有了长足的进步和发展,中兴通讯在大视频运维中也适时采用了机器学习方法来提取历史巡检数据中蕴含的故障特征,并构建集成预测模型来提升大视频运维的精度和效率。从当前

Backblaze 数据集的测试情况来看,也取得了较好的效果。

在当前工作的基础上,我们后续将进一步提升人工智能在大视频运维中的落地效果,包括采用半监督学习来提高模型的数据利用率,采用迁移学习来加速模型在新局点的训练部署进度,使用强化学习来优化大视频运维的策略和流程等。

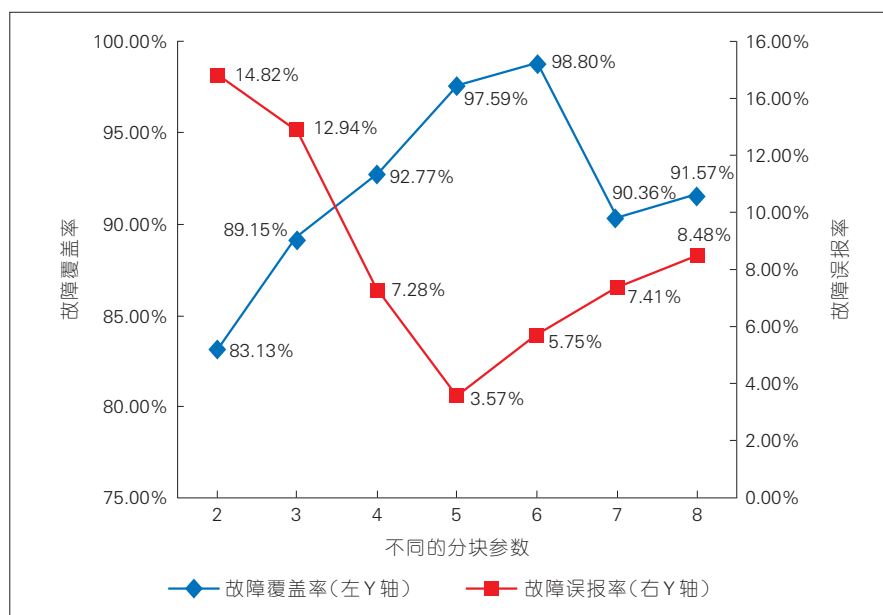
3 结束语

信息通信技术(ICT)时代,无论对于运营商网络还是业务系统的运维支撑,都需要加速与人工智能技术的落地实践,提供高度自动化和智能化的运维解决方案。人工智能、机器学习技术在大视频运维的智能化提升重点体现在运维模式从被动式事后分析转为积极主动预测、分析及决策。随着人工智能技术的加速发展,

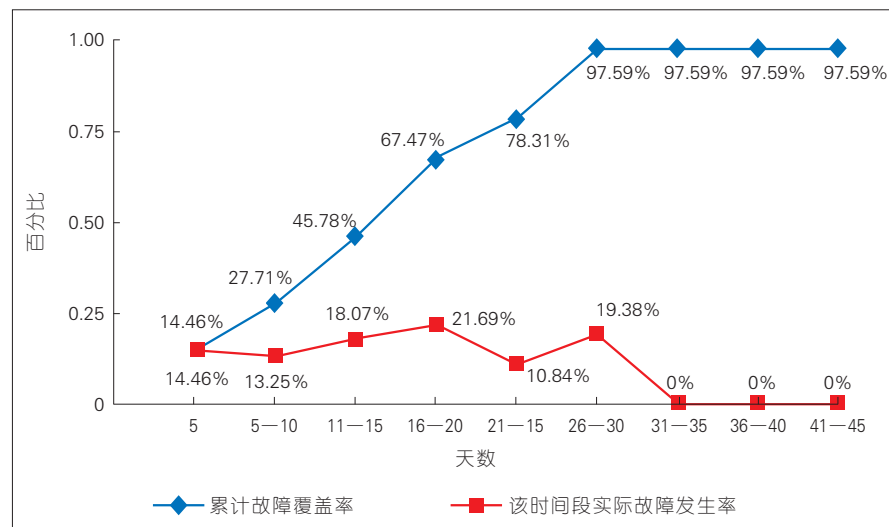
▼表 1 采用的 SMART 基础属性表

ID	SMART 属性名称
1	重映射扇区计数
2	重映射扇区计数(原始值)
3	当前待映射扇区计数
4	当前待映射扇区计数(原始值)
5	底层数据读取错误率
6	主轴起旋时间
7	寻道错误率
8	通电时间
9	(磁头)高飞写入
10	温度(摄氏度)
11	无法校正的错误
12	硬件 ECC 校正

ECC:错误检查和校正技术
SMART:自我检测分析和报告技术



▲图 6 随旋转森林分块系数变化的故障覆盖率和误报率 (预警提前量为 30 天情况)



▲ 图7 预警提前天数分布累计(最低误报率情况下)

大视频运维与人工智能技术的结合会越来越紧密,大视频运维技术将朝着更加智能化的方向演进,实现更加自动化和精准的故障预测和排查,主动发现业务系统中的故障或薄弱环节并加以修复。在实现智能运维基础上,通过对视频业务使用者的行为分析、家庭及用户画像等一系列的建模分析,充分挖掘海量数据的价值,衍生出新的业务形态,实现智能化的运营系统,为运营商创造新的商机。

参考文献

- [1] 黄珂,李锐,姜春鹤.基于大数据的视频体验保障[J].中兴通讯技术(简讯),2017(3):22-25
- [2] 基于机器学习的智能运维[EB/OL].(2017-04-22)[2017-06-25].<https://zhuanlan.zhihu.com/p/26216857>

- [3] TANG L, LI T, PERNG C S. LogSig: Generating System Events from Raw Textual Logs[C]//In Proceedings ACM International Conference on Information and knowledge Management. UK:ACM, 2011:785-794
- [4] TANG L, LI T. LogTree: A Framework for Generating System Events from Raw Textual Logs[C]//In Proceedings of IEEE International Conference on Data Mining (ICDM). USA: IEEE, 2010:491-500
- [5] LUO C, LOU J G, LIN Q W, et al. Correlating Events with Time Series for Incident Diagnosis[C]//In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM, 2014: 1583-1592
- [6] HAMERLY G, ELKAN C. Bayesian Approaches to Failure Prediction for Disk Drives[EB/OL].(2017-06-28). <http://cseweb.ucsd.edu/~elkan/smart.pdf> ICML
- [7] ECKART B, CHEN X, HE X, et al. Failure Prediction Models for Proactive Fault Tolerance within Storage Systems[C]//

Modeling, Analysis and Simulation of Computers and Telecommunication Systems 2008, IEEE International Symposium on. USA: IEEE, 2008. DOI:10.1109/MASCOT.2008.4770560

- [8] ECKART B, CHEN X, HE X, et al. Failure Prediction Models for Proactive Fault Tolerance within Storage Systems[J]. IEEE International Symposium on Modeling, 2009, 1(3):1-8. DOI:10.1109/MASCOT.2008.4770560
- [9] Hard Drive Reliability Statistics [EB/OL]. [2017-06-28].<https://www.backblaze.com/b2/hard-drive-test-data.html>

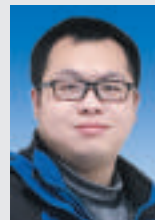
作者简介



屠要峰,中兴通讯股份有限公司云计算及IT研究院副院长;长期从事电信业务及云计算产品的研发工作,主要研究方向云计算、大数据及人工智能;已发表论文10余篇。



吉锋,中兴通讯股份有限公司云计算及IT研究院项目经理;先后从事过IPTV/OTT/CDN、移动互联网相关的产品研发、标准&技术的研究工作,目前研究方向为大数据与人工智能;已发表论文3篇。



文韬,中兴通讯股份有限公司云计算及IT研究院高级系统工程师;研究方向为机器学习与智能运维、推荐系统的结合。

基于 3D CNN 的大规模视频手势识别研究

Large-Scale Video-Based Gesture Recognition Using 3D CNN Model

苗启广/MIAO Qiguang
李宇楠/LI Yunan
徐昕/XU Xin

(西安电子科技大学, 陕西 西安 710071)
(Xidian University, Xi'an 710071, China)

手势是一种交流的形式,它指的是利用人的肢体动作来说明其意图或态度的行为。由于在视频监测控制、标志语言理解、虚拟现实和人机交互等领域有着巨大的应用前景,越来越多的研究人员开始研究手势识别算法,以实现将人类手势解释给机器的目标。

手势识别最早期的研究是从 20 世纪 80 年代开始,是一种典型的涉及到各方面知识的研究。为了对人体动作的含义加以识别,研究人员先后使用了大量不同种类的方法。早期的大部分方法都是基于复杂的人工制作特征。Stamer 和 Pentl^[1]首先利用隐马尔科夫模型(HMM)进行美国手语的相关研究;Elmezain^[2]等利用 HMM 模型对手势的动态轨迹进行了识别;Sgouropoulos 等人^[3]将神经网络方法和 HMM 方法结合使用,可提高

收稿日期: 2017-05-21

网络出版日期: 2017-07-07

基金项目: 国家自然科学基金(61472302、U1404620、61672409); 模式识别国家重点实验室开放课题基金资助(201600031); 中央高校基本科研业务费专项资金(JB150317); 陕西省自然科学基金(2010JM8027); 航空科学基金(2015ZC31005)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0009-005

摘要: 提出了一种基于三维卷积神经网络(CNN)的大规模视频手势识别算法。首先,为了获得统一尺度的输入数据,在时域上对所有输入视频进行了归一化处理得到 32 帧的输入视频;然后,为了从不同的角度描述手势特征,通过真彩(RGB)视频数据生成了光流视频,并将 RGB 视频和光流视频分别通过 C3D 模型(一个 3D CNN 模型)提取特征,并通过特征连接的方式加以融合输入到支持向量机(SVM)分类器来提高识别性能。该方法在 Chalearn LAP 独立手势数据集(IsoGD)的验证集上达到了 46.70% 的准确率。

关键词: 手势识别; 三维卷积神经网络; 光流; SVM

Abstract: In this paper, an effective 3D convolutional neural network(CNN)-based method for large-scale gesture recognition is proposed. To obtain compact and uniform data for training and feature extracting, the inputs are unified into 32-frame videos. To describe features of gesture in different aspects, the optical flow data from red, green, blue (RGB) videos are generated. After that, the spatiotemporal features of RGB and optical flow data are extracted with the C3D model (a 3D CNN model) respectively and blended together in the next stage to boost the performance. Finally, the classes are predicted with a linear support vector machine (SVM) classifier. Our proposed method achieves 46.70% accuracy on the validation set of ChalearnLAP Isolated Gesture Dataset (IsoGD).

Keywords: gesture recognition; 3D CNN; optical flow; SVM

动态手势的识别效果,并且该方法具有光照鲁棒性。Wan 等人将尺度不变特征变换(SIFT)特征加以扩充,得到 3D 增强动作 SIFT (EMoSIFT)^[4] 和 3D 稀疏动作 SIFT (SMoSIFT)^[5],并通过稀疏关键点混合特征(MFSK)^[6]来进行手势识别。随着近来深度学习技术的蓬勃发展,利用深度神经网络实现手势识别受到了研究者的广泛关注,且相对于传统手工特征方法,在识别率上也取得了重大突破。Karpathy 等人^[7]利用卷积神经网络(CNN)来实现手势行为识别;

Simonyan 和 Zisserman^[8]利用一个双流 CNN 网络同时提取手势视频中的时域和空域的特征;Tran 等人^[9]提出了一个 3D CNN 模型——C3D 模型,解决了基于视频的手势识别需要同时处理时域和空域的特征这一问题。

在文章中,我们提出了一种基于同源数据融合的大规模的手势识别方法。首先,我们通过对数据分布特征的分析,将所有帧数不一的视频进行预处理,获得统一的帧数为 32 帧的视频;随后,我们由真彩(RGB)视频生成了光流视频,以进一步提取动

作信息,同时避免表演者服饰、肤色等因素的干扰;我们再利用上文提到的C3D模型,对RGB数据和光流数据同时提取空域和时域上的特征信息,并将这些特征加以融合,最终通过支持向量机(SVM)分类器来获得分类结果,整个流程如图1所示。

1 基于C3D模型的视频手势识别

1.1 预处理

一般的CNN由于其中全连接层的限制,都要求输入数据具有相同的大小。因此我们首先需要对数据进行归一化处理,即统一的帧数,各帧相同的宽和高。为了尽可能地获取代表手势含义的特征,我们采取一种数据驱动的策略来实现这一过程,即通过对数据分布情况的分析来确定归一化的方式。首先,我们分析了实验数据集——Chalearn LAP IsoGD Database(简称IsoGD数据集),该数据集由Wan等人^[10]建立,它源自于Guyon等人^[11]建立的ChaLearn手势数据集(CGD)。IsoGD数据集包含了47 933个独立的视频,每个视频包含一个手势,这些手势被分为249类,它被用于2016年首届Chalearn LAP大规模手势识别竞赛,其详细信息如表1所示。由于IsoGD数据集中,每个视频的宽度和高度都是一致的,因此需要归一化处理的主要是时域,即帧数信息。

▼表1 Chalearn LAP IsoGD数据集详细信息

子集	类别数	手势数量	表演者数量
训练集	249	35 878	17
校验集	249	5 784	2
测试集	249	6 271	2

如图2所示,由于在数据集中,一些类别的手势看起来非常相似,因此在处理视频成统一帧数时,就需要在保留动作的运动路径信息和降低视频的空间占用之间进行折衷。在分析了35 878个训练集的视频的帧数后,我们发现:尽管视频的帧数从1~405帧各不相同,但是大部分视频的帧数在29~39之间,其中33帧的视频数量最多,达1 202个。为了便于处理,我们选择32作为视频的基准帧数,将所有视频统一至32帧。帧数大于32的视频需进行采样,而帧数小于32的视频则通过复制按一定比例选出的帧进行插值。通过这样的预处理方式,超过98%的视频至少每3帧进行了1次采样,大部分的运动路径信息得以保留。

1.2 光流特征提取

光流,是视觉场中的一种目标、表面和边缘的表征运动的模型,它是由观察者和场景之间的相对运动产生的。在文章中,我们通过RGB视频提取光流特征,一方面用于提取动作路径信息,另一方面也去除了背景、表演者肤色等与动作无关的信息。我们通过Brox等人^[12]提出的基于亮

度恒常性、梯度恒常性和时空平滑约束假设的能量方程来计算光流特征,该能量方程可表述为:

$$E(u,v) = E_{Data} + \alpha E_{Smooth} \quad (1)$$

其中, $\alpha > 0$ 是一个正则化参数, E_{Data} 可表述为:

$$E_{Data}(u,v) = \int_{\Omega} \psi(|\Delta I(x)|^2) + \gamma |\Delta G(x)|^2 dx \quad (2)$$

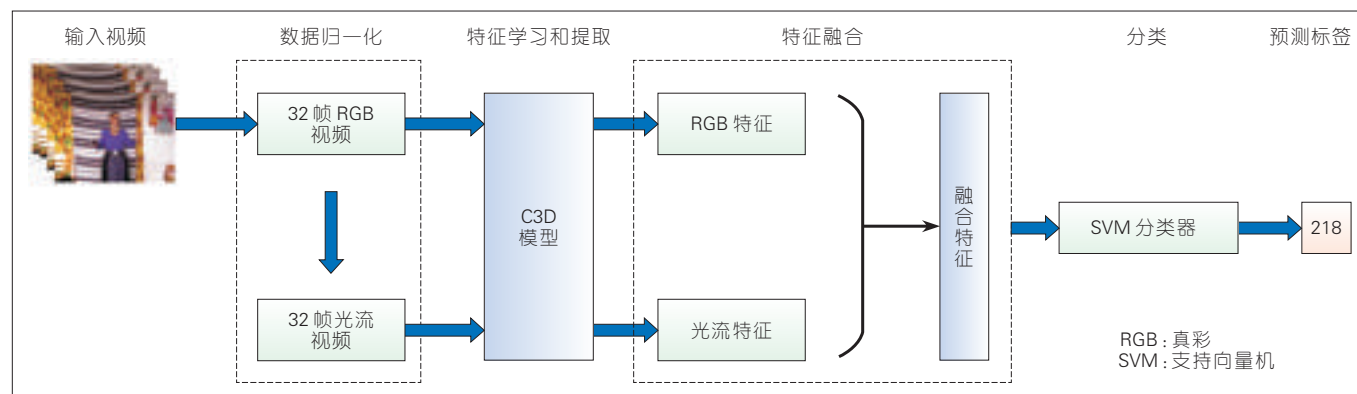
其中, γ 是用于平衡两者的权重系数, ΔI 和 ΔG 是视频两帧之间的灰度和梯度增量, $\psi(s^2)$ 用来增强能量方程的鲁棒性, Ω 为积分区间,即整个视频, E_{Smooth} 可表述为:

$$E_{Smooth}(u,v) = \int_{\Omega} \psi(|\nabla_3 u|^2 + |\nabla_3 v|^2) dx \quad (3)$$

∇_3 表示时空平滑约束假设中的时空梯度。通过利用拉格朗日方程和数值近似来最小化该能量函数,获得最终的光流结果。显然,光流数据更加关注运动信息,能够把运动无关的信息全部去除。

1.3 特征提取模型

如前所述,基于深度神经网络的特征提取由于能够更好地体现数据本身的信息,且不像手工特征那样需要研究者具备大量领域相关信息,因



▲图1 基于C3D模型和同源数据融合的大规模手势识别算法流程



▲图2 相似手势举例

而受到了研究者的青睐。文中所述需要提取的特征关注的手势是在视频中,所以解决手势识别任务更多的是依靠提取到的时序特征。因此,我们通过一种三维CNN——C3D模型来实现视频手势特征的自动提取。与二维的CNN相比,三维的CNN针对视频帧序列图像集合,并不仅仅是把视频划分成为帧集合,再用多通道输出到多个图像,而是将卷积核应用到时域,时空域的卷积核特性相互结合,更好地获取视频的特征。

如图3所示,C3D模型包括8个卷积层、5个池化层、2个全连接层来学习特征,和1个softmax层来提供预测的类别。8个卷积层的卷积核个数分别是64、128、256、256、512、512、512和512,卷积核的最佳大小是 $3 \times 3 \times 3$ 。通过对视频的时空卷积,可以获得在不同尺度上的特征图。在1次或2次卷积操作之后,通过1次池化操作,来对特征进行降采样,以获得更具全局性的特征。在文中,第2~5层的池化层的卷积核大小是 $2 \times 2 \times 2$,而第1个池化层的卷积核大小是 $1 \times 2 \times 2$,以保证在网络中时域信息能够得到最大程度上的保留。在经过多次卷积和池化操作之后,特征图被抽象成一个4 096维的特征向量,

用来标记样本的分类信息。

1.4 融合方案

Tran等人结合3个不同的网络提取出的特征来提高准确率。这给了我们灵感:使用特征融合的方法可以提高识别能力。通过实验我们发现:由于RGB视频和光流视频在描述特征的方式并不相同,因此直接将两个视频简单融合,反而不利于正确率的提升。相反,因为特征是视频的抽象,对于C3D提取出的特征向量,可以很好地阐述手势的特点。因此,我们选择了特征级融合。这样做的另一个优势是特征都是相同维度的,统一的格式有助于正确、有效地融合。为了保证两种数据的信息能够同时保留,我们选择通过将两种特征拼接得到高维特征的方式来实现融合。

2 手势识别实验结果和分析

由于目前IsoGD数据集的测试集部分的标签尚未公开,所以文中提到的所有实验和比较都是在该数据集的验证集上进行的。

2.1 实验环境

文中所提到的神经网络训练和特征提取在配有Intel Core i7-6700

CPU @ 3.40 GHz、16 GB内存和Nvidia GeForce GTX TITAN X图形处理器(GPU)的PC上实现,C3D模型依托Linux Ubuntu 16.04长期支持版本(LTS)系统和caffe框架实现32帧视频的生成,特征融合和SVM分类则在Windows 7系统(64 bit)上通过Matlab R2012b软件实现。

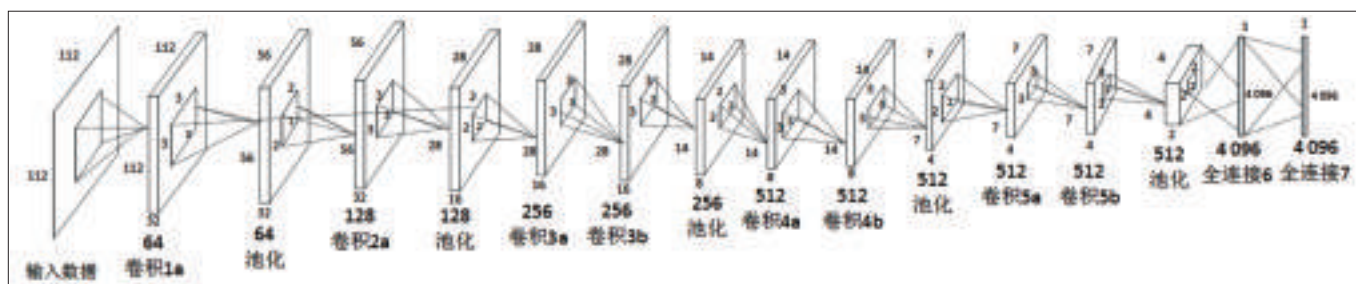
2.2 训练过程

考虑到训练一个较深的网络是很耗时的,特别是在IsoGD这样的大型数据集上,因此我们首先通过Sport-1M(最大的视频分类的标准数据集,包含110万个运动视频,共487类)预训练模型,使其能够适应视频动作分类的应用场景,随后再在实验所需的IsoGD数据集上调参。我们通过随机梯度下降法(SGD)来训练网络:首先将数据打乱,以减少数据排列信息对训练的干扰,在每一次训练的迭代过程中,有10个视频输入网络,网络初始学习率设为0.0001,并且在每5 000次迭代后以10%的比例下降,在10万次迭代后训练停止。

2.3 迭代的影响

作为一个基于学习的方法,迭代次数对分类结果有很大的影响。因此,在这个部分,我们分别在RGB和光流数据的输入上测试不同迭代次数的影响。识别率和损失函数值的变化情况如图4所示。

在训练过程的早期,网络的学习能力较强,损失函数值下降很快,在经过约3万次迭代后,RGB和光流数据的损失函数值都趋于稳定。最后,



▲图3 C3D网络结构

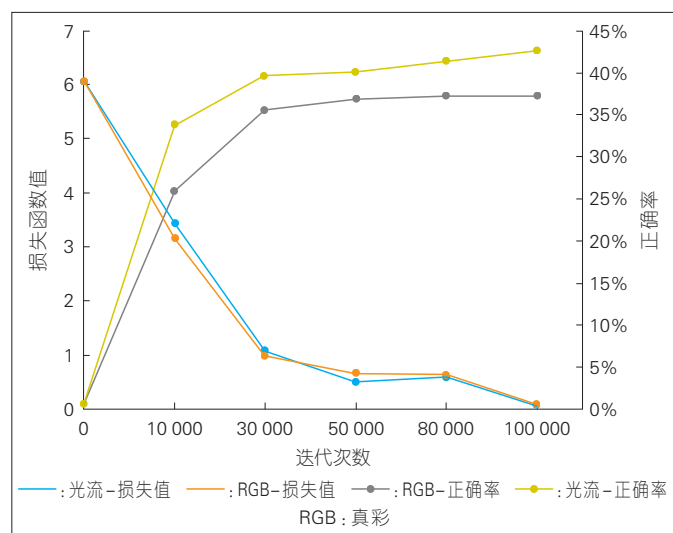


图4
RGB与光流数据
损失函数值与正确
率随迭代次数变化
关系

在10万次迭代之后,损失函数值非常接近于0,这时网络已经从训练数据中学到了足够多的东西。另一方面,识别率显示出类似的趋势:正确率在训练早期上升很快。同时,不同数据的特征间的关系也在这个阶段展现出来。在1万次迭代之后,光流数据的优势开始显现出来,而且直到最后一次迭代,光流数据的准确率一直比RGB的高5%左右。

2.4 预处理效果

在本节中,我们验证预处理,即32帧归一化策略的效果。我们对比了输入视频分别为16帧和32帧的结果,这两种输入都是10万次迭代后的结果。

如图5所示,通过我们对输入数据的分析,32帧的归一化策略取得显著的效果,无论是在RGB还是光流数据上,相比16帧的输入,两种数据的32帧的输入都提高了约4%的识别率。这证明更多关于运动路径的信息有助于分辨不同的手势,从而在很大程度上提高了识别率。

2.5 融合结果

我们通过融合RGB和光流特征来提高性能。融合结果和单独RGB、光流数据的比较如图6所示。

相比于任何一种单一特征,融合

策略对识别性能的提升都是显著的。融合特征相比单独RGB特征将正确率提升了将近10%,相比单独的光流特征,融合特征也有近5%的提升。这证明了特征融合的策略是行

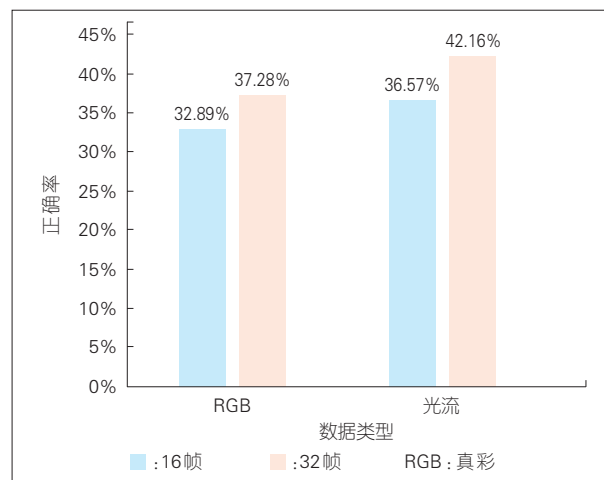


图5
32帧归一化预处理效果

之有效的。

2.6 和传统方法的比较

在本节中,我们将我们的方法与Wan等人提出的基于手工制作特征的方法进行对比,从图7中可以看出:CNN在对图片或视频的特征提取方面展示出了极大的优势,我们的方法大概将识别率提高了30%。

2.7 最终结果对比

我们的方法和大赛中前3名的方

法,以及大赛的基准方法在校验集上的结果对比如图8所示。结果显示:相对于使用手工制作特征的基准方法,我们基于深度学习的方法在特征提取上具有更好的性能。此外,使用光流数据进行去背景处理,使我们的识别率更进一步提升,相对于大赛中的方法,加入光流数据使得准确率提升了4%。

3 结束语

我们提出了一种基于RGB和光流数据及三维CNN的手势识别方法。输入的数据首先被统一成32帧的视频,以便更好地保存动作路径信息;然后,我们通过RGB数据生成了光流数据来去除视频中与手势无关的因素;接着,RGB和光流视频的特征被C3D模型分别提取并加以融合来提高识别性能;之后,我们使用

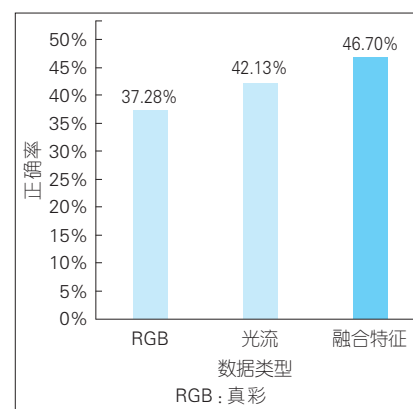


图6 RGB与光流数据融合结果

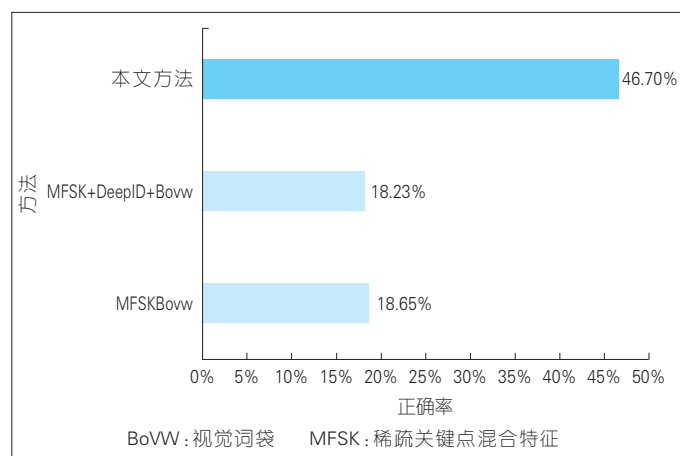


图7
基于 C3D 的手势
识别算法与传统
手工特征方法结果
比较

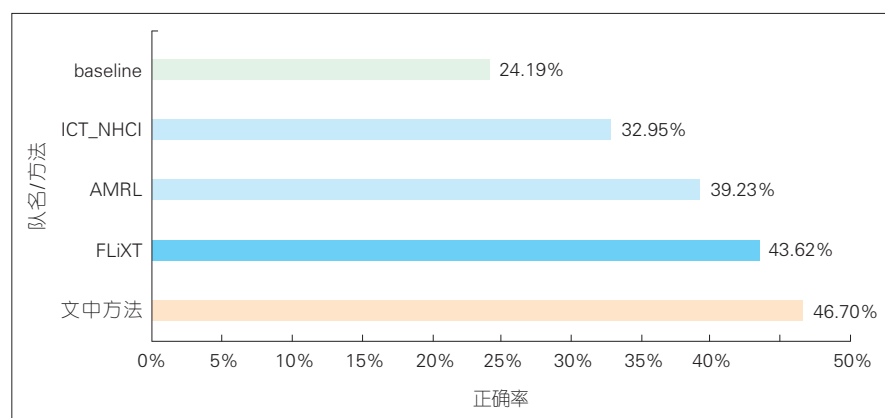


图8 文中方法与 Chalearn LAP 大规模手势识别竞赛方法结果比较

SVM 分类器进行最终分类。实验证明:我们的策略是有效的,而且我们的方法优于其他现有的技术。相较于 Chalearn LAP 大规模手势识别竞赛中的方法及传统手工特征方法,我们方法的识别正确率有了较大提升。

然而,仍然存在着很多因素影响识别率。由于运动信息还不足以区分那些差别细微的类别,还需要学习更多复杂的特征来解决这些问题。同时,还有很多其他的深度学习网络结构,如深度置信网络,在目标识别方面展示出了很大的优势。这些网络结构在视频手势识别方面的使用还值得更多的研究。

致谢

本文的部分实验由西安电子科技大学计算机学院硕士研究生田宽和范莹莹完成,在此对他们表示衷心

感谢!

参考文献

- [1] STARNER T, PENTL A. Visual Recognition of American Sign Language Using Hidden Markov Models[J]. International Workshop on Automatic Face&Gesture Recognition, 1995(2):189-194
- [2] ELMZAIN M, HAMADI A, MICHAELIS B. Hand Trajectory-Based Gesture Spotting and Recognition using HMM[C]//The 16th IEEE International Conference on Image Processing (ICIP). USA:IEEE,2009:3577-3580. DOI: 10.1109/ICIP.2009.5414322
- [3] SGOUROPOULOS K, STERGIOPOULOU E, PAPAMARKOS N.A Dynamic Gesture and Posture Recognition system[J]. Journal of Intelligent&Robotic Systems, 2013(1):1-14
- [4] WAN J, RUAN Q, LI W, et al. One-Shot Learning Gesture Recognition from RGB-D Data Using Bag of Features[J]. Journal of Machine Learning Research, 2013, 14(1): 2549-2582
- [5] WAN J, RUAN Q, LI W, et al. 3D SMO-SIFT: Three-Dimensional Sparse Motion Scale Invariant Feature Transform for Activity Recognition from RGB-D Videos[J]. Journal of Electronic Imaging, 2014, 23(2): 023017-023017
- [6] WAN J, RUAN Q, LI W, et al. One-Shot Learning Gesture Recognition from RGB-D

Data Using Bag of Features[J]. Journal of Machine Learning Research, 2013, 14(1): 2549-2582

- [7] KARPATY A, TODERICI G, SHETTY S, et al. Large-Scale Video Classification with Convolutional Neural Networks[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. USA: IEEE, 2014: 1725-1732
- [8] SIMONYAN K, ZISSERMAN A. Two-Stream Convolutional Networks for Action Recognition in Videos[C]//Advances in Neural Information Processing Systems. Canada: NIPS, 2014: 568-576
- [9] TRAN D, BOURDEV L, FERGUS R, et al. Learning Spatiotemporal Features with 3d Convolutional Networks[C]//Proceedings of the IEEE International Conference on Computer Vision. USA: IEEE, 2015: 4489-4497
- [10] WAN J, ZHAO Y, ZHOU S, et al. Chalearn Looking at People RGB-D Isolated and Continuous Datasets for Gesture Recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. USA:IEEE, 2016: 56-64
- [11] GUYON I, ATHITSOS V, JANGYODSUK P, et al. The ChaLearn Gesture Dataset (CGD 2011)[J]. Machine Vision and Applications, 2014, 25(8): 1929-1951
- [12] BROS T, BRUHN A, PAPENBERG N, et al. High Accuracy Optical Flow Estimation Based on a Theory for Warping[J]. Computer Vision-ECCV 2004, 2004(3024): 25-36. DOI: 10.1007/978-3-540-24673-2_3

作者简介



苗启广,西安电子科技大学计算机学院教授;主要从事计算机视觉、机器学习、大数据分析方面的研究;主持在研或完成国家核高基重大专项、国家重点研发计划、国家自然科学基金、省自然科学基金、国防预研、国防“863”和“新世纪优秀人才支持计划”项目20余项;在重要期刊上发表论文100余篇。



李宇楠,西安电子科技大学计算机学院博士研究生;主要从事计算机视觉、深度学习方面的研究。



徐昕,西安电子科技大学计算机学院硕士研究生;主要从事计算机视觉、深度学习方面的研究。

基于深度学习的多目标跟踪算法研究

Multiple Object Tracking Algorithm Based on Deep Learning

陆平/LU Ping
邓硕/DENG Shuo
李伟华/LI Weihua

(中兴通讯股份有限公司, 广东 深圳
518052)
(ZTE Corporation, Shenzhen 518052,
China)

自从深度学习技术出现以来, 计算机视觉领域得到了快速发展, 深度学习技术最先用于图像分类问题。近年来, 基于深度学习的多目标跟踪算法也取得了一定的突破。多目标跟踪是计算机视觉领域一个非常具有挑战性的研究方向, 且有着十分广泛的现实应用场景, 例如: 智能视频监控控制、异常行为分析、移动机器人研究等。传统的多目标跟踪算法往往由于目标检测效果较差, 导致跟踪效果不佳; 而基于深度学习的检测器可以获得较好的目标检测效果, 进而提高目标跟踪的准确度。

因此, 文章中我们着重研究了基于深度学习的多目标跟踪算法。首先通过 GoogLeNet^[1]+长短期记忆网络(LSTM)^[2]模型进行目标检测, 以获得准确的目标检测结果。在此基础上, 提出了直接根据目标检测的特征图对检测目标进行深度特征的提取的方法, 深度特征相比于尺度不变特征变换(SIFT)^[3]等传统特征可以更准确地反应检测目标的外观特征, 因此可以提高目标跟踪算法的准确性。此外, 还在基于马尔科夫蒙特卡洛

收稿日期: 2017-05-20
网络出版日期: 2017-07-06

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0014-006

摘要: 提出了一种基于深度学习的多目标跟踪算法。首先, 通过 GoogLeNet+长短期记忆网络(LSTM)模型进行目标检测, 以获得准确的目标检测结果; 其次, 直接根据目标检测的特征图对检测目标进行深度特征的提取, 深度特征相比于传统特征可以更准确地反映检测目标的外观特征, 因此可以有效提高跟踪的准确性。此外, 还在传统数据驱动马尔科夫蒙特卡洛(DDMCMC)算法的基础上, 提出了层次的数据驱动马尔科夫蒙特卡洛(HDDMCMC)算法, 可以进一步提高多目标跟踪的准确性。实验结果证明了所提出算法的有效性。

关键词: 多目标跟踪; 深度学习; 目标检测; MCMC 算法

Abstract: In this paper, a multi-target tracking algorithm based on deep learning is proposed. Firstly, GoogLeNet + long short-term memory (LSTM) model is used to obtain accurate object detection results. Secondly, the feature map of object detection is directly used to extract the deep feature for tracking. Compared with the traditional feature, the deep feature can reflect the appearance of objects more accurately, which could improve the tracking accuracy effectively. What's more, based on the traditional Data Driven Markov Chain Monte Carlo (DDMCMC) algorithm, the Hierarchical Data Driven Markov Chain Monte Carlo (HDDMCMC) algorithm is proposed to further improve the tracking accuracy. The experiment results prove the effectiveness of our algorithm.

Keywords: multiple object tracking; deep learning; object detection; MCMC algorithm

(MCMC)算法的多目标跟踪算法的基础上, 提出了层次的数据驱动马尔科夫蒙特卡洛(HDDMCMC)算法。

1 基于深度学习的多目标跟踪算法的研究现状

1.1 传统多目标跟踪算法

MCMC 算法是一种经典的多目标跟踪算法, Yu 等人在 MCMC 算法的基础上提出了数据驱动的马尔科夫蒙特卡洛(DDMCMC)算法^[4]。在使用该算法获得检测数据后, 我们按照传统的 MCMC 算法, 对当前轨迹中的检测数据进行位置和长相特征的衡量, 来

计算当前的后验概率大小, 然后在迭代过程中不断进行不同状态之间的转移来进行寻找全局最优结果。

Tang 等人提出了一种基于图分割的多目标跟踪算法^[5], 通过在时间和空间上对边界框进行聚类来进行轨迹匹配。Tang 等人在用该算法解决轨迹匹配这一最优化问题时, 提出了一种基于 KL (Kernighan-Lin) 算法的近似解法, 运用该算法求得的结果与剪枝法求得的结果相比, 准确率略微下降, 但是运算速度有较大提高。

1.2 基于深度学习的多目标跟踪算法

Bing 等人提出了一种联合学习

卷积神经网络(CNN)特征和时域约束度量的模型,通过这个训练模型结合相应的损失函数可以构建出轨迹亲和力模型,再通过传统的图匹配方法将所有的轨迹进行联合,利用softassign算法求出轨迹匹配最优解,得到最终结果。

Fengwei等人提出了一种基于深度学习的目标检测外观特征,将这种特征应用到多目标跟踪后,获得了较高的跟踪准确度,但需要大量的训练数据和时间,且无法达到实时跟踪的效果。

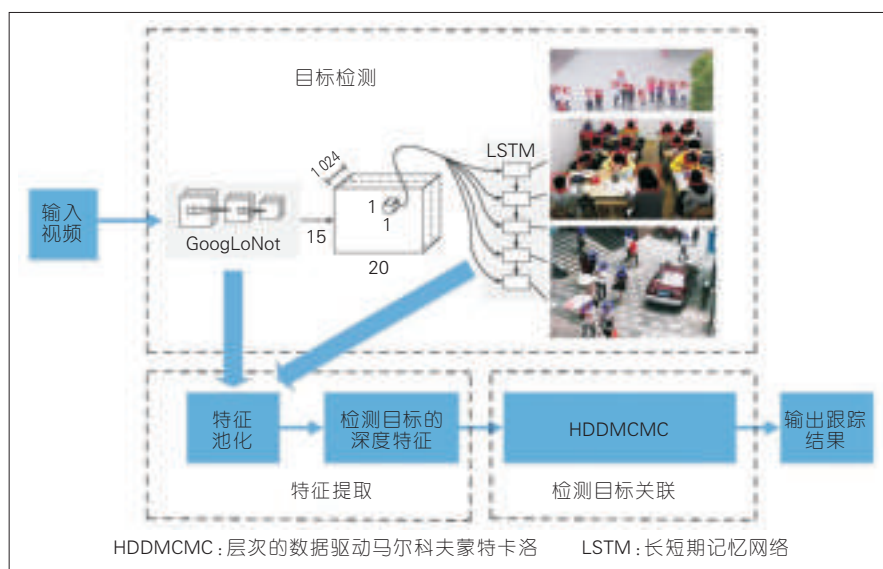
基于深度学习的多目标跟踪算法的跟踪准确度相对于传统多目标跟踪算法有了较大提升,但仍需要大量的训练数据和时间,且实时性不足。针对这一问题,我们在基于深度学习的目标检测算法的基础上,提出一种可直接利用目标检测的特征图得到深度特征的多目标跟踪算法,并基于HDDMCMC算法得到跟踪结果。

2 基于目标检测特征图的多目标跟踪算法框架

我们提出的算法框架如图1所示。算法整体上可以分为目标检测、特征提取、检测目标关联这3个模块。对于输入视频,首先通过GoogLeNet+LSTM模型进行目标检测,根据检测模块中的特征图,对检测目标提取深度特征,利用深度特征可以计算外观相似度,进而通过DDMCMC算法对检测目标进行匹配,形成完整的轨迹,得到跟踪结果。基于目标检测特征图的特征提取模块和基于HDDMCMC算法的检测目标关联模块是该算法的主要创新点。

3 基于GoogLeNet+LSTM的目标检测

目标检测是基于数据联合的多目标跟踪算法的基础,针对在复杂场景下目标过小和目标遮挡等问题,我们利用GoogLeNet+LSTM这一框架进行目标检测。对于输入视频,我们首



▲图1 基于深度学习的多目标跟踪算法框架

先使用GoogLeNet进行卷积,在最后一层得到 $1 \times 1024 \times 15 \times 20$ 的特征图阵列,把它进行转置成为 300×1024 的特征图阵列。每个1024维的向量对应原图中 139×139 的区域。

在利用GoogLeNet进行卷积以后可以得到 300×1024 的特征图阵列,然后通过LSTM子模块并行处理每一个1024维向量。对于每一个输出的隐状态,经过两个不同的全连接层:一个直接输出框的位置和宽高,一个再经过softmax层输出这个框的置信度。LSTM子模块共有5个这样的LSTM单元,即对于每个输入预测5个可能的框和对应的置信度,在训练中使得框的位置集中在感知区域中心的 64×64 的位置,置信度是从高到低排的。

每个1024维的向量经过LSTM阵列处理以后,可以得到对应原图中 64×64 小块的5个检测框以及对应的置信度。因为LSTM是并行处理的,所以生成的对 64×64 区域的检测结果是重叠的。最后框处理子模块需要对整个视频帧的所有检测框进行筛选,然后再通过给定一个阈值去掉置信度低的框,并给出最终检测结果。具体流程是:对于已经确定的框,如果一个待选的框与它有相交,

则去除这个框,限制一个已经确定的框至多去除一个待选框。在上述匹配的时候代价用 (m, d) 衡量, m 表示评估的二者是否相交,取值 $\{0,1\}$, d 衡量评估的两个框之间曼哈顿距离。 m 的重要性大于 d ,即对两种匹配方案得到的结果,先比较 m 大小,如果不能得出结论,再比较 d 的大小。我们使用匈牙利算法来寻找代价最小的匹配方式,假设筛选的置信度阈值设为0.5,那么则去掉那些置信度低于0.5的框。

为有效对目标检测模型进行训练,我们采用了如下训练方法:首先在LSTM子模块得到较多待选框,但存在检测错误或误差。错误或误差的情况有3种:

- (1)把不是人头尖的地方框出;
- (2)预测框位置和真值框位置的差异;
- (3)对同一个目标产生了多个预测框。

对于情况(1),可通过赋予待选框较低的置信度来杜绝;对于情况(2),需要修正待选框和与之匹配的真值框之间的误差;对于情况(3),可通过给同一个目标后生成的预测框赋予较低的置信度来杜绝。

模型训练时的损失函数如公式

(1)所示:

$$L(G, C, f) = \alpha \sum_{i=1}^{|G|} l_{pos}(\hat{b}_{pos}^i, \hat{b}_{pos}^{f(i)}) + \sum_{j=1}^C l_c(\hat{b}_{pos}^j, \{f^{-1}(j)\}) \quad (1)$$

其中, G 代表框的真值, C 代表待选框, f 代表匹配算法, \hat{b}_{pos}^i 代表真值框里面的第 i 个框, \hat{b}_{pos}^j 代表待选框里面第 j 个框, l_{pos} 代表两者之间的曼哈顿距离, l_c 是交叉熵损失, 也就是对应网络里面的 softmax 损失。这个损失函数的前一项代表待选框和与之匹配的真值框的位置误差, 后一项代表待选框的置信度, α 调整这两种损失之间的平衡。

匹配算法为匈牙利算法, 其中用的比较函数如公式(2)所示:

$$\Delta(b_i, \hat{b}_j) = (o_{ij}, r_j, d_{ij}) \quad (2)$$

其中, o_{ij} 取值为 {0,1}, 若待选框的中心落在真值框中, 则为 0, 否则为 1; r_j 为该待选框生成的序号, 目的是在匹配的时候, 偏好先生成置信度比较高的框, 故匹配同一个目标时, 排序越靠前代价越低; d_{ij} 是两个框之间的距离, 即距离误差。

4 基于目标检测特征图的深度特征提取

在检测模块中得到对应框的位置时, 得到的是人头的位置, 按一定比例放大可以覆盖全身。得到行人的框位置和大小后, 我们进一步利用 GoogLeNet 的最后一层卷积层得到的特征图阵列去提取特征。我们使用的方法是快速-区域 CNN (RCNN)^[6] 中提到的感兴趣区域 (ROI) 池化, 即根据特征图阵列相对于输入图片的缩小比例, 把原图中的感兴趣区域在特征图阵列中对应的感兴趣区域池化成一个 1 024 维的向量。其中缩小比例为 32 倍, 即得到框的位置按 32 倍比例缩小, 并且为了减少背景影响, 最大化池化这一池化类型更加有效。通过将检测模块中的特征图进行池化, 可得到对检测模块中检测到

的每个目标的深度特征, 因为每个特征都是高度抽象的, 可以很好地表征目标的外观特征。

文中, 我们所提出算法的重要特点就是利用目标检测中的特征图进行池化来得到目标跟踪所需要的深度特征, 而不需要重新进行训练, 因此可以在不牺牲目标跟踪算法的实时性的前提下, 提高跟踪准确性。

5 基于 HDDMCMC 算法的目标检测关联

5.1 传统 MCMC 算法

为了对检测数据进行最优关联, 可利用 MCMC 算法进行建模。传统 MCMC 算法的计算过程为: 在迭代过程中, 均匀随机选取一个转移动作 (包括产生、消失、融合、分裂、扩展、收缩和交换等), 把当前的行人轨迹按照此转移动作进行转换, 即从状态 w 转移到状态 w' 。在此情况下, 可以计算得到转移前后的后验概率 $\pi(w)$ 和 $\pi(w')$, 以及转移概率 $q(w|w', D)$ 和 $q(w'w, D)$, 从而可以计算得出此时的接收概率 $A(w', w)$ 。将此接收概率与从均匀 0~1 分布中随机抽取的值 u 进行比较, 如果出现 $A(w', w) > u$, 则接收当前这个转移, 使得当前状态转移为 w' , 即 $w = w'$, 否则不接受此转移。为了得到最大后验概率, 我们再将转移后的状态与当前的最优状态 w^* 进行比较, 如果转移后的状态优于当前的最优状态, 则对最优状态进行更新, 即 $w^* = w$ 。

在传统 MCMC 算法的基础上, 我们提出了 HDDMCMC 算法, 即把传统的单层 MCMC 算法分为段内 MCMC 算法和段间 MCMC 算法。如图 2 所示, 假设当前处理的时间段是 $[t+2T, t+3T-1]$, 我们首先对此段执行段内 MCMC 算法并得到相应的段内行人轨迹。在此之前, $[t, t+T-1]$ 段和 $[t+T, t+2T-1]$ 段内的 MCMC 算法已执行完成, 并执行了两者之间的段间 MCMC 算法, 从而得到对应时间

帧内的行人轨迹; 然后把其中已走出当前区域的行人轨迹选择出来, 再对 $[t, t+2T-1]$ 段和 $[t+2T, t+3T-1]$ 段再执行段间 MCMC 算法, 得到当前的行人轨迹。HDDMCMC 算法即按照此种方式不断执行, 直到视频序列结束。

5.2 段内 MCMC 算法

在多目标跟踪算法中, 考虑到运动的稳定性和连续性 (即同一个目标在前后帧视频数据中, 外观特征不会发生剧烈变化), 在段内 MCMC 算法中, 使用第 4 节中的深度特征对目标轨迹的相似度进行度量。

可以将每个检测目标看成一个节点, 以段内时间 $[t, t+T]$ 来进行叙述。假设视频帧 t 内的节点集合为 $N_t = \{N^t(1), N^t(2), N^t(3), \dots, N^t(i), \dots, N^t(N_t-1), N^t(N_t)\}$, 后验概率 $P(w|D)$ 的设定如公式(3)所示:

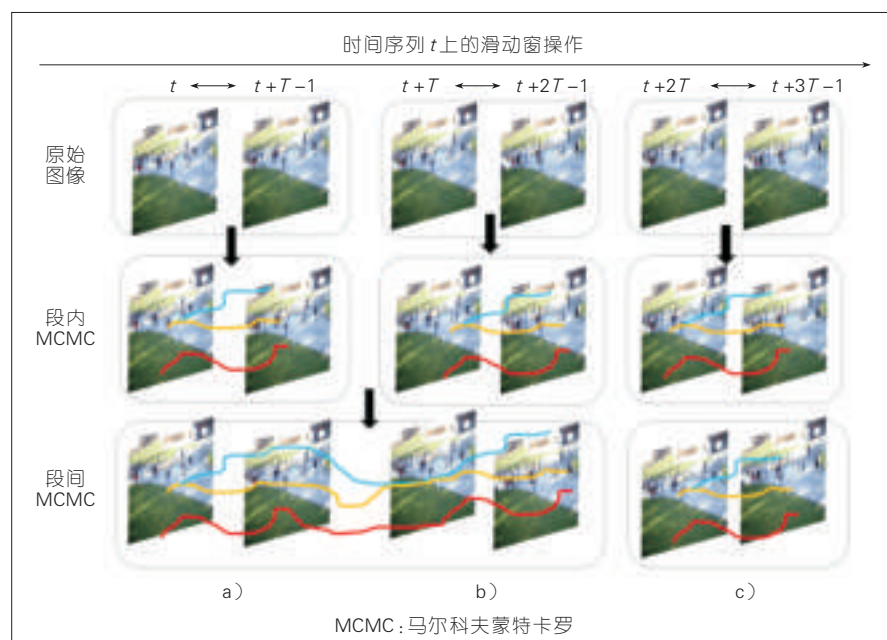
$$P(w|D) = P(\{\tau_1, \tau_2, \tau_3, \dots, \tau_k\} | D) = \left(\prod_{k=1}^K \prod_{n=1}^{N_k-1} P(N^{n+1}(k) | N^n(k)) \right) \times \left(\prod_{k=1}^K e^{-\frac{1}{l_k}} \right) \times (e^{-|\tau_0|}) \quad (3)$$

其中, $N^n(k)$ 和 $N^{n+1}(k)$ 分别代表第 k 条行人轨迹中第 $n+1$ 个和第 n 个节点, $P(N^{n+1}(k) | N^n(k))$ 即为两节点的相似度, 可用两节点深度特征的夹角余弦值来进行计算。公式第 2 项中的 l_k 代表的是不同轨迹的长度, 保证行人轨迹的完整; 公式第 3 项中的 $|\tau_0|$ 代表的是虚警的个数, 保证虚警率较低。

图 3 为 MCMC 算法中 7 种操作: 产生和消失这两个转移动作如图 3a) 所示, 其中, 左边黑色的点代表虚警集合中的点, 经过产生操作后转化为右边红色点所代表的行人轨迹, 消失操作为反向顺序; 融合和分裂这两个转移动作如图 3b) 所示; 扩展和收缩这两个转移动作如图 3c) 所示, 交换和转移动作如图 3d) 所示, 此操作前后互为逆操作。

5.3 段间 MCMC 算法

段间 MCMC 算法使用的数据主



▲图2 HDDMCMC 算法

要是段内 MCMC 算法生成的目标轨迹。在段间 MCMC 算法中,主要采取的转移动作包括融合和分裂操作以及交换操作。因为,在经过段内 MCMC 算法之后,生成许多较为可靠的目标轨迹。此时若存在同一目标轨迹断裂的情况,就是由于检测数据不稳定,对应目标碰撞或者遮挡导致对应的漏检帧数过多等原因造成的。因此,段间 MCMC 的目的是将两个时间段的目标轨迹数据做进一步的数据联合。当前状态下,后验概率项更新为公式(4)所示。

$$P(w|D) = P(\{\tau_1, \tau_2, \tau_3, \dots, \tau_K\} | D) = \left(\prod_{k=1}^K \prod_{n=1}^{N_k-1} P(N^{n+1}(k) | N^n(k)) \right) \times \left(\prod_{k=1}^K e^{-\frac{1}{l_k}} \right) \quad (4)$$

在公式(4)中,我们不再考虑虚警因素,因为此处主要是对目标轨迹的划分操作。由于已有之前较为可靠的两个时间段内的目标轨迹数据作为辅助,因此可以在更加宽松的条件下进行抽样工作。对于融合操作,此时允许的时间间隔设定为 $t_{gap} = 10$,且两个目标的轨迹段与段之间连接处的帧差不能超过6。对于概率中的标准差的设定,此时为

$3\sigma = \text{size}(d_i^t)$,即允许在理想位置周围的变化幅度是一个当前目标的大小。对于分裂操作和交换操作,抽样的时间节点在段与段连接处的时间

帧为左右各6帧时间以内。通过这种方式,可使得不同状态之间进行转移的单位是之前已生成的较为完整的目标轨迹片段。

在当前的段间 MCMC 算法结合之后,把已走出视频场景的目标移除当前数据集,此时的当前数据集假设为 τ' 。当下一个段内 MCMC 得到轨迹之后,通过段间 MCMC 算法与 τ' 进行匹配,即继续在之前目标数据的基础之上,结合当前的目标数据,做进一步的数据联合来优化。整个算法按照这样的滑动方式不断进行。

6 实验结果及分析

6.1 MOT2015 训练数据库的实验结果

本小节中,我们介绍的是 MOT2015 训练数据库^[8]中的实验结果,此数据库总共包括11个视频,有固定视角、运动视角等场景,同时又有俯视视角、平时视角等场景,由于场景变化多样,不同目标之间的接触

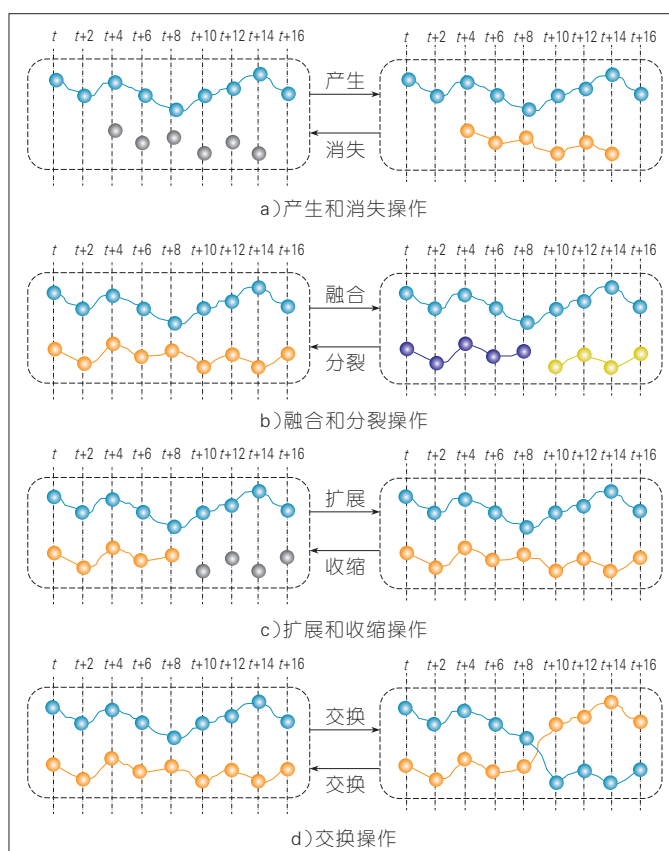


图3

7种转移操作示意

和碰撞较多,挑战性较大。

我们使用的衡量指标^[9]包括多目标跟踪的准确度(MOTA)、多目标跟踪的精确度(MOTP)、漏检数目(FN)、虚警数目(FP)等,其中MOTA指标是用来衡量整体跟踪准确度,最具代表性。实验结果如表1所示,其中MOTA和MOTP是11个数据库的平均值,FN、FP和ID Sw是11个数据库的总和。

文章中,我们所提出算法的主要创新点为基于目标检测特征图的特征提取模块和基于层次的数据驱动马尔科夫蒙特卡洛算法的检测目标关联模块,为了验证这两个创新点的有效性,我们分别用4种方法进行了实验:

(1) SIFT+MCMC,即用SIFT特征结合传统MCMC算法。

(2) SIFT+HDDMCMC,即用传统SIFT特征结合文中所提出的HDDMCMC算法。

(3) CNN+MCMC,即用文中提出的通过目标检测特征图提取的深度特征结合传统MCMC算法。

(4) CNN+HDDMCMC,即用通过目标检测特征图提取的深度特征结合HDDMCMC算法,也即文中我们提出算法的完整版。

实验结果如表1所示,对比SIFT+MCMC和SIFT+HDDMCMC可知:我们提出的基于层次的HDDMCMC算法相比于传统的MCMC算法可以在一定程度上提升跟踪效果;对比SIFT+MCMC和CNN+MCMC可知:通过用目标检测特征图提取的深度特征来代替传统的SIFT特征,可以在很大程度上提升跟踪效果;CNN+HDDMCMC算法,即文中我们提出算法的完整版的各项检测指标都优于其他方法,更进一步证明了所提出算法的有效性。

图4和图5是两张跟踪效果的对比图(为便于观察,只显示了要说明的区域)。图4中第1行是使用SIFT+HDDMCMC的实验结果,第2行是使用CNN+HDDMCMC的实验结果,每1

行的3张图片代表视频中间的连续3帧(前后2帧之间间隔1帧)。第1行的第3张图片在红色箭头标识处出现了跟踪错误,这是由于目标之间的遮挡且用SIFT特征求目标之间的外观相似度时并不十分准确而产生的;第2行图片中,由于深度特征能更准确地反应检测目标的外观特征,没有出现错误。

图5中第1行是使用CNN+MCMC的实验结果,第2行是使用CNN+HDDMCMC的实验结果。每1行的第2张图片由于目标之间的遮挡发生了漏检,导致了轨迹断裂,使得第1行中的第3张图片在红色箭头标识处出现了错误;而在第2行图片中,由于此处使用了HDDMCMC算法,通过段间MCMC算法对两段轨迹进行了匹配,从而避免了错误。

基于此数据库的对比实验可知:文中所提出的算法对于生成稳定的目标轨迹具有十分重要的作用。

6.2 MOT2015测试数据库实验结果及其对比

本小节中,我们介绍的是MOT2015测试数据库中的实验结果,此数据库同样包括11个视频,我们将文中所提出算法的实验结果与其他算法进行了对比,比较方法包括如下3种:

(1) 结构化支持向量机算法(LP_SVM)^[10],此算法是Shaofei等人提出的,主要思想是使用最小网络流模式,此模型的参数通过结构化支持向量机训练得到。

(2) 近时多目标跟踪算法(NOMT)^[11],此算法是由Wongun等人

▼表1 MOT2015训练数据库实验结果

	MOTA	MOTP	FN	FP
SIFT+MCMC	36.2	73.6	27 872	14 967
SIFT+HDDMCMC	39.8	73.6	25 688	13 576
CNN+MCMC	45.3	73.7	20 854	11 378
CNN+HDDMCMC	47.6	73.8	19 638	8 895

CNN:卷积神经网络
FN:漏检数目
FP:虚警数目

HDDMCMC:层次的数据驱动马尔科夫蒙特卡洛
MCMC:马尔科夫蒙特卡洛

MOTA:多目标跟踪的准确度
MOTP:多目标跟踪的精确度
SIFT:尺度不变特征变换

图4▶
是否使用深度特征的跟踪效果对比

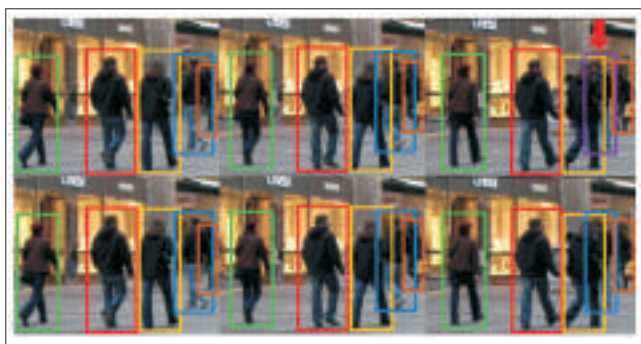
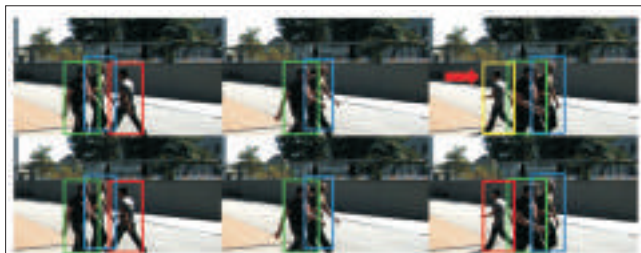


图5▶
是否使用HDDMCMC的跟踪效果对比



提出的,主要思想是通过建立一个聚合本地光流描述子来计算检测目标之间的相似度,实现了接近实时的多目标跟踪。

(3) 子卷积神经网络算法(MDP_SubCNN)^[12],此算法是由Yu等人提出的,主要思想是通过学习的策略来建立检测目标之间的相似度矩阵,并通过马尔科夫决策过程建模。

实验结果如表2所示,其中Our表示的是文中我们提出的算法,与其他法对比可知:本算法的MOTA指标最高,整体效果最好;FN值优于其他算法;FP值相对于其他算法较低;MOTP值只略逊于MDP_SubCNN算法,总的来说,我们提出算法的效果较为理想。

7 结束语

多目标跟踪作为计算机视觉领域中的一个经典问题,有着十分广泛的应用前景。近年来,随着深度学习技术的发展,基于深度学习的多目标跟踪算法取得了一定的突破,获得了高于传统多目标跟踪算法的跟踪准确度。在文章中,我们提出了一种新的基于深度学习的多目标跟踪算法,通过用GoogLeNet+LSTM框架进行目标检测的特征图对检测目标深度特征的提取,以及对传统MCMC算法的改进,有效地提高目标跟踪算法的准确性和实时性。最后对文中所提出的算法进行了实验验证,并与一些相关的多目标跟踪算法进行了比较,得到了较为理想的结果。

致谢

本研究得到了上海交通大学彭瑾龙、林巍晓老师的指导与帮助,对他们谨致谢意!

参考文献

- [1] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA:IEEE, 2015: 1-9. DOI: 10.1109/CVPR.2015.7298594

▼表2 MOT2015 测试数据库实验结果

	MOTA	MOTP	FN	FP
LP_S SVM	25.2	71.7	36 932	8 369
NOMT	33.7	71.9	32 547	7 762
MDP_SubCNN	47.5	74.2	22 969	8 631
Our	48.2	74.0	21 318	9 385

CNN: 卷积神经网络
FN: 漏检数目
FP: 虚警数目
HDDMCMC: 层次的数据驱动马尔科夫蒙特卡洛

LP_S SVM: 结构化支持向量机算法
MCMC: 马尔科夫蒙特卡洛
MDP_SubCNN: 子卷积神经网络算法

MOTA: 多目标跟踪的准确度
MOTP: 多目标跟踪的精确度
SIFT: 尺度不变特征变换

- [2] HOCHREITER S, SCHMIDUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780
- [3] LOWER D G. Distinctive Image Features from Scale-Invariant Key Points[J]. International Journal of Computer Vision, 2004, 60(2): 91-110
- [4] LOWE G D. Object Recognition from Local Scale-Invariant Features[C]// ICCV '99 Proceedings of the International Conference on Computer Vision. USA:IEEE, 1999
- [5] OH S, RUSSELL S, SASTRY S. Markov Chain Monte Carlo Data Association for General Multiple-Target Tracking Problems[C]// Decision and Control, 2004. CDC. 43rd IEEE Conference on. USA:IEEE, 2004
- [6] GIRSHICK R. Fast R-CNN[EB/OL].(2015-09-27)[2017-06-25]. <https://arxiv.org/abs/1504.08083>
- [7] YU Q, MEDIONI G, COHEN I. Multiple Target Tracking Using Spatio-Temporal Markov Chain Monte Carlo data association [C]// Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007. USA:IEEE, 2007. DOI: 10.1109/CVPR.2007.382991
- [8] Multiple Object Tracking Benchmark[EB/OL]. [2017-06-25]. <https://motchallenge.net/>
- [9] BERNARDIN K, STIEFELHAGEN R. Evaluating Multiple Object Tracking Performance: the CLEAR MOT Metrics[J]. EURASIP Journal on Image and Video Processing, 2008, 2008(1): 1-10
- [10] WANG S F, FOWLKES C C. Learning Optimal Parameters for Multi-target Tracking[EB/OL].(2016-10-05)[2017-06-25]. <https://arxiv.org/abs/1610.01394>
- [11] CHOI W. Near-Online Multi-Target Tracking with Aggregated Local Flow Descriptor[C]//Proceedings of the IEEE International Conference on Computer Vision. USA:IEEE, 2015: 3029-3037
- [12] XIANG Y, ALAHI A, SAVARESE S. Learning to Track: Online Multi-Object Tracking by Decision Making[C]//Proceedings of the IEEE International Conference on Computer Vision. USA: IEEE, 2015: 4705-4713. DOI: 10.1109/ICCV.2015.534
- [13] MILAN A, LEAL T L, SCHINDLER K, et al. Joint Tracking and Segmentation of Multiple Targets[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA:IEEE, 2015. DOI: 10.1109/CVPR.2015.7299178
- [14] YOON J H, YANG M H, LIM JONGWOO, et al. Bayesian Multi-Object tracking Using Motion Context from Multiple Objects[C]// Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on. USA: IEEE, 2015. DOI: 10.1109/WACV.2015.12
- [15] MCLAUGHLIN N, RINCON J M D, MILLER P. Enhancing Linear Programming with Motion Modeling for Multi-target Tracking [C]// Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on. USA:IEEE, 2015
- [16] XIANG Y, ALAHI A, SAVARESE S, et al. Learning to Track: Online Multi-Object Tracking by Decision Making[C]// Computer Vision (ICCV), 2015 IEEE Conference on. USA:IEEE, 2015. DOI: 10.1109/ICCV.2015.534
- [17] BEWLEY A, GE Z, OTT L, et al. Simple Online and Realtime Tracking[C]//Image Processing (ICIP), 2016 IEEE International Conference on. USA: IEEE, 2016: 3464-3468. DOI: 10.1109/ICIP.2016.7533003

作者简介



陆平,中兴通讯股份有限公司云计算及政企业务产品总经理,北京邮电大学和南京邮电大学兼职教授;主要从事云计算、大数据、增强现实、基于多媒体服务技术等方面的研究;主持国家、省部级基金项目10余项;获得省部级科技进步奖多项;拥有20余项发明专利,发表论文15篇,出版专著2部。



邓硕,中兴通讯股份有限公司高级工程师;研究方向为图像处理、视频智能分析等技术,主要从事基于深度学习的图像智能分析技术预研与开发;参与多项中兴通讯产学研合作项目及国家、省部级基金项目。



李伟华,中兴通讯股份有限公司高级工程师;主要研究领域包括人工智能、图像及视频分析等方面;拥有20余项发明专利,承担标准Editor席位并输出多篇国际和国内标准。

车辆属性识别及跨场景标注方案

A Solution for Vehicle Attributes Recognition and Cross-Scene Annotation

董振江/DONG Zhenjiang¹
高燕/GAO Yan¹
吴文熙/WU Wenxi²

(1. 中兴通讯股份有限公司, 广东 深圳 518057)

(2. 中山大学, 广东 广州 510006)

(1. ZTE Corporation, Shenzhen 518057, China)

(2. Sun Yat-Sen University, Guangzhou 510006, China)

智能交通^[1-3]领域中牵涉到很多计算机视觉的任务,需要对海量的视频、图像数据进行处理、分析。在基于车辆视频图像的分析任务中,传统技术不仅效率较低,耗费巨大的人力、物力、财力,而且欠缺对数据的感知表达能力,没有充分利用数据、挖掘数据特性,以致分析精度不够高。引入车辆属性识别的方法,可以增强算法对目标车辆特征的理解和表达能力,大大提高分析任务的可靠性和效率。

随着计算机视觉和机器学习技术的发展,以它们为技术基础的新一代智能分析技术,在目标检测、识别、检索等任务中表现优越,能够极大地减少人力成本,提高监测控制的效率,有巨大的应用潜力。

自2012年ImageNet^[4]国际大规模视觉识别比赛以来,业界掀起了一波基于计算机视觉和深度学习的识别技术研究和应用热潮。深度学习是一门大数据驱动的技术,我们通常认

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0020-005

摘要: 提出采用一种基于深度学习的识别方法,来辅助获取训练模型中所需要的车辆属性标签。该方法首先利用海报图像构建大规模车辆属性数据集(SYSU-ZTE-CARS),训练基于卷积神经网络(CNN)的识别模型,再将模型迁移到监测控制场景中进行标注测试,间接获取属性标签。采用CNN+softmax分类器的结构作为基本框架,引入细粒度识别技术以进一步优化识别性能。利用5种常用车辆属性进行测试,实验结果表明:所提出的方案不仅在SYSU-ZTE-CARS数据集上的识别精度高,而且在监测控制场景下的标注结果也很可靠。

关键词: 精细化属性识别; CNN; 深度学习; 计算机视觉

Abstract: The deep learning-based solution is proposed to obtain the vehicle attributes data in recognizer training. Firstly, with poster images collected, a large scale vehicle attributes dataset—SYSU-ZTE-CARS, is built to train the convolutional neural network(CNN)-based recognition model. Then the model is moved to the monitor and control scene to mark test, and access attribute indirectly. CNN + softmax classifier is used as the basic framework, and the fine-grained identification is also used to further optimize the recognition performance. The experimental results on recognition of five different attributes show that our solution is not only with high accuracy in SYSU-ZTE-CAR dataset, but also reliable in auto-annotation under real-world surveillance.

Keywords: fine-grained attributes recognition; CNN; deep learning; computer vision

为:提供训练的数据量越大,所训练得到的模型性能越好;而在视频监测控制的应用场景下,海量的视频图像数据会源源不断地产生,因此将深度学习技术应用在视频监测控制领域,恰能发挥其先天优势以及视频数据的作用。

基于深度学习的属性识别技术,通常以卷积神经网络(CNN)为基本工具。CNN模型参数的学习,目前使用最广泛也是最稳定的方法是有监督的训练策略。这种策略需要使用大量车辆图片及其属性标签作为训练样本。目前面临最大的挑战是:在现实监测控制场景下的车辆属性标

签很难获取。虽然有些属性标签可以通过人工标注的方式获取,如颜色、视角、车类型、品牌等,但是这种方式同时也带来了新的问题,如:颜色易受光照影响,在极端光照环境下,人眼无法正确判别车辆颜色;品牌往往需要通过车标来确定,一旦车标模糊、被遮挡、不可见,则需要通过人眼来判别车辆品牌。

另一方面,我们注意到:在互联网上有大量带有属性标签的车辆海报图片。为了获取监测控制场景下的车辆属性标签,我们考虑了一种跨场景自动标注的技术方案。该方案中,我们先利用带有属性标签的车辆

收稿日期: 2017-05-25
网络出版日期: 2017-07-04

海报图片训练属性识别模型,再将此模型迁移到现实的监测控制场景中,自动标注出车辆属性。

1 车辆属性识别的相关工作

1.1 车辆属性识别的研究

目前,对于车辆属性的研究工作和应用都已经相当广泛,主要集中于视角识别、颜色识别、车类型识别、品牌识别等几个方面。

在视角识别方面,学术界主要在KITTI基准测试平台^[4]上进行。KITTI上的研究绝大部分是基于CNN的,其识别效果精度都非常高,有些优秀的工作如多尺度CNN(MSCNN)^[5]、循环滚动卷积(RRC)^[6]等,以及一些利用三维信息的工作,能够在直接预测车辆偏转角度的具体数值的任务上有可靠的性能表现。

在颜色识别方面,研究难点在于:光照影响对颜色判断的干扰特别严重。目前主要的研究多是基于强判别性区域辅助的方式,具体实现手段主要包括利用关键点、利用局部分割等。在颜色特征表达上,主要有基于阈值分割、颜色直方图和CNN等。

在车类型识别方面,可以分成两大类:基于非视觉的方法和基于视觉的方法。基于非视觉的方法,业界通常采用如雷达、超声波、红外线等传感器信号开展研究应用;基于视觉的方法则通过通用摄像头采集视频图像数据,再通过计算机视觉技术进行识别。目前的研究主要有基于直方图特征、低维投影特征和稀疏表示特征等传统方法。

在品牌与子品牌识别方面,研究成果同样也可以分为传统方法和深度学习方法两大类。传统方法主要是通过传统特征工程的手段人工设计特征,如Curvelet变换、方向梯度直方图(HOG)特征、分层梯度直方图(PHOG)特征、边缘方向直方图(EOH)特征、Gabor小波等,也有基于融合特征进行稀疏编码的解决方

案。深度学习方法,主要是基于CNN特征对属性进行判别的,也有一些针对细粒度识别的工作,如Part-based R-CNN模型、基于分割图辅助的CNN模型、双线性CNN模型等。

总体来说,基于传统方法的车辆识别技术都有特征设计过程反复、特征表达能力鲁棒性差、模型泛化能力差的缺点;而基于深度学习的车辆属性识别技术,则具有强大的数据描述能力,且在泛化性、识别精度、稳定性上都比传统方法更胜一筹。

1.2 CNN网络的研究

属性识别往往通过分类技术实现,而分类问题是作为基于CNN的视觉识别中一项最基本任务,其性能最主要体现在CNN网络结构之上,CNN网络的发展直接就能够反映分类精度变化。

在大规模视觉识别挑战(ILSVRC)2012比赛中,AlexNet的成功再次掀起了学者们对CNN的研究热潮。CNN结构上的发展体现在两个主要的维度上——深度和广度。在深度上,从8层的AlexNet,到16层的VGG^[7]、24层的GoogLeNet,再到152层的ResNet。在Cifar数据集上,ResNet还成功实现了极深网络,层数达到1 001层。在2016年的ImageNet比赛中,商汤科技-香港中文大学团队更是使用了1 200层的极深网络。在广度上,GoogLeNet^[8]首先设计出Inception模块,在同一层网络里使用了更多子网络。

研究表明:CNN的层数越深,网络表达能力越强。文章中,我们基于ResNet系列的网络结构开展实验,同时考虑到任务难度,也采用了轻量级网络,如AlexNet;兼顾精度和测试效率,我们也采用了中等深度网络,如VGG16。

1.3 相关数据集的情况

关于车辆属性图片数据集,目前比较有代表性的主要有CompCars^[10]、

BoxCars^[10]、城镇交通运输监测控制系统数据集(UTS)^[11]。

CompCars数据集由香港中文大学所提出,是目前学术界规模最大的车辆精细化属性数据集。数据集分为两部分:一部分来自于各大汽车门户网站中的图片;另一部分来自于道路监测控制视频截取出的图片。第1部分收集161个汽车品牌、1 687个子品牌汽车,子品牌下还细分了汽车的生产年份,其中整车图片有136 727张,汽车部分细节图片有27 618张,这些图片涵盖车辆的5种视角、12种类型;第2部分有50 000张汽车正面的监测控制图像,同样标注了车类型、品牌和子品牌,除此之外,第2部分数据还标注了10种颜色。BoxCars数据集也是汽车海报图片,标注内容和方式与CompCars相似,但其规模远不如CompCars。UTS数据集收集的是监测控制场景下的原始图片,是目前在该场景下标注内容最丰富、最贴近现实应用场景的图片。UTS数据集共3 600张图片,涵盖了6种场景,在每种场景内包含多辆车,它的标注内容包括车辆位置、5种视角、9种颜色和6种常见车类型。

虽然UTS数据集非常贴近现实应用场景,但是由于标注量较小、属性丰富性欠缺等原因,使得无法训练得到泛化能力较强的属性识别模型。然而,CompCars的属性标签也未与现实场景完全吻合,如CompCars并没有覆盖到如“货车”“公交车”等常见车类型,其第2部分数据虽然来自于卡口监测控制场景,但视角单一。为了弥补CompCars和UTS数据集的不足,文中我们自行构建了大规模的车辆属性数据集SYSU-ZTE-CARS。

2 基于深度学习的车辆属性识别方案

2.1 基于CNN的基本技术框架

基于深度学习的车辆属性识别方案,主要以CNN作为基本工具。

CNN是一种参数可学习的模型,通常是利用反向传播(BP)算法对网络参数进行微调更新,有着数据拟合能力强,训练简便,支持端到端的训练和测试的特点。在属性识别技术方案(如图1所示)中,CNN的作用是作为特征提取器,对输入的图片/视频帧进行特征表达,得到其具有较强判别性的特征向量;再利用分类器对特征向量进行分类,以此实现对图片/视频帧中车辆属性的识别。

文中所提出的方案选用了 AlexNet^[12]、VGG16^[7]、ResNet^[13]等经典CNN模型,利用有监督学习的方法,训练了多种属性识别网络模型,对不同属性的识别性能进行了测试评估和分析。

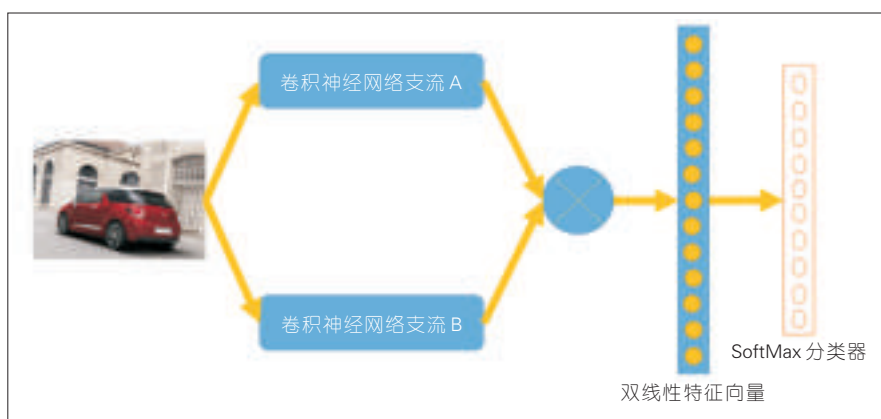
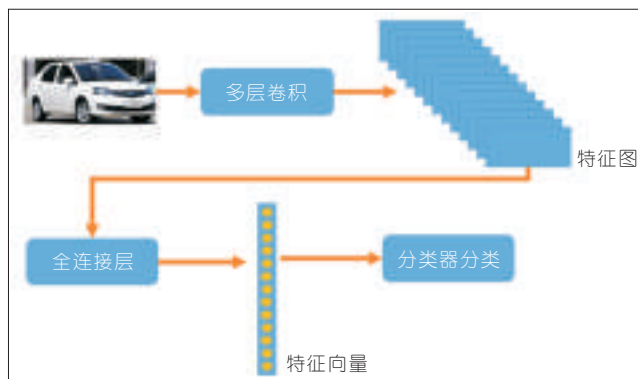
2.2 基于细粒度的识别优化方案

属性识别,本质上是类内区分。对于类内判别性较强的属性,如视角、颜色、车类型,直接采用“CNN+分类器”的方法已能够获得比较好的识别效果;但对于类内区分度不明显的属性,如品牌和子品牌,我们通常需要使用基于细粒度的识别技术来提高识别的精确度。文中,我们采用双线性CNN(B-CNN)^[14]的方法,使用B-CNN网络框架(如图2所示),增强特征向量对于细粒度分类识别的表达能力。

B-CNN延续了一般物体分类的做法,即通过CNN提取特征,然后用向量表示特征,最后用分类器分类向量。一般物体分类的做法是用全连接层将特征图转化为一条特征向量,作者认为这种做法导致特征向量的辨别程度不高,难以辨识相似度高的物体。在工作中,使用的是双流CNN,它分别提取了两个特征图,并且将在这两个特征图每个位置算外积,然后用sum-pooling池化,最后再拉直成一条特征向量^[14]。用这个特征向量进行细粒度的分类识别,效果有显著提升。

假设图1某个区域*i*经过两个

图1
基于CNN的属性识别
基本技术框架



▲ 图2 B-CNN原理

CNN对应的特征图分别是矩阵 A ($C \times M$ 维)和 B ($C \times N$ 维),其中 M 和 N 是两个特征图的深度, C 可以是两位的数对,代表宽和高。

那么此时两矩阵相乘 $A^T \times B$ 的结果是矩阵 P_i ($M \times N$ 维),则

$$\Phi(I) = \sum_{i \in I} P_i \quad (1)$$

将 $\Phi(I)$ 拉直成一条向量 L ($MN \times 1$ 维),即图2中的双线性特征向量。对这个特征向量做标准化处理后便可用作分类。在对细粒度属性表达能力上,该方法所得到的特征向量比全连接层输出的特征向量的更强。

2.3 车辆属性识别演示系统

文中,我们研究并实现了一个车辆属性识别演示系统,此原型系统的设计主要参考MVC框架,将界面、业务逻辑、数据三者分层,在修改业务逻辑时只需要修改控制层的部件,在修改用户界面(UI)时,只需要对定义

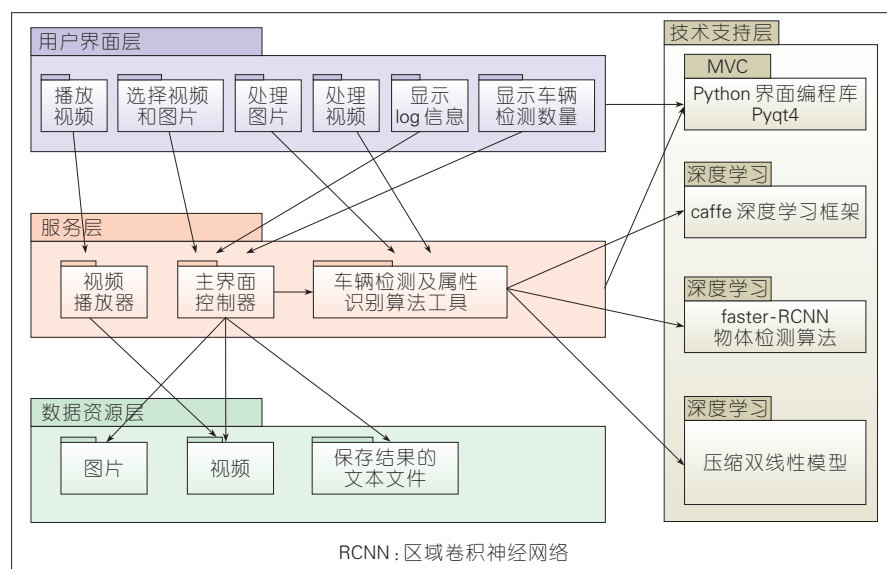
UI的xml文件做出修改,不需要改变控制层的代码,降低了代码的耦合性,增加重用性。系统的总体架构如图3所示:最上层是UI层,中间是业务逻辑层,下面是数据资源层,而右边则是该系统的技术支持层。

演示系统使用python+pyqt4开发,界面简洁。系统主要功能是对输入的图像和视频做车辆检测和属性识别处理,并输出结果。系统的深度学习框架采用Caffe^[15],Caffe是采用C++语言开发的高清晰、高可读性、高效的深度学习框架,支持多语言接口。

3 车辆属性识别实验结果与分析

3.1 SYSU-ZTE-CARS数据集

我们使用有监督的训练方法训练属性识别模型,该方法需要训练样本,即图片及其属性类别标签。因此,我们参考CompCars数据集的构建



▲ 图3 演示系统架构

方式和经验,利用网络爬虫技术,从互联网获取了大量车辆图片,并对所有图片进行了属性标注,构建了用于训练属性识别模型的大规模数据集 SYSU-ZTE-CARS,图4是该数据集的样例图片。

SYSU-ZTE-CARS 数据集共有 169 610 张的车辆图片,共标注了 5 种常用车辆属性,其属性标签标注情况如表1所示。其中,属性标签包括正前、正后、正侧、前侧和后侧共 5 种车辆视角;包括黑、白、红、蓝、黄、绿、紫、棕、灰、银共 10 种颜色;包括“MPV”“SUV”“掀背车”“轿车”等 15 种类型,以及 215 种品牌和 1 609 种子品牌。

3.2 车辆属性识别实验结果

在算法的性能测试实验中,我们初步采用单任务学习框架,在 SYSU-ZTE-CARS 数据集上对属性识别算法性能进行评估,如表2所示。所谓的单任务学习框架,即对每一种属性,分别训练一个 CNN 模型,各属性识别模型之间的参数不共享。

评测指标我们采用 Top-1 精度评价策略,即把模型预测结果按置信度从高到低排序,只考虑置信度最高的预测结果是否正确。

实验结果表明:对于简单的属

性,如视角、颜色、车类型,采用轻量级网络 (AlexNet) 或中等深度网络 (VGG16),已经有比较可靠的识别精度。采用极深网络 (ResNet 系列) 虽然对识别精度有小幅提升,但同时也增加了模型的参数规模,增大了计算量。细粒度识别技术对于简单属性的识别精度提升不但帮助不大,反而增加了模型运算开销。对于较困难的属性,如品牌,轻量级网络的特征判别性表达能力不足,增加网络深度,可以明显提高识别精度。采用细粒度识别优化后,其识别精度有小幅上升。对于困难的属性,如子品牌,仅仅增加网络深度并不能获得较好的识别性能,而须要通过细粒度的识别技术来增强特征的可判别性,才能达到可靠的识别精度。



▲ 图4 SYSU-ZTE-CARS 数据集样例

▼ 表1 SYSU-ZTE-CARS 数据集属性标签标注情况

属性	视角	颜色	车类型	品牌	子品牌
种类	5	10	15	215	1 609

▼ 表2 各种属性在 SYSU-ZTE-CARS 数据集上的初步测评效果

方法	top-1 精度 / %				
	视角	颜色	车类型	品牌	子品牌
AlexNet+Softmax	87.10	89.01	88.30	76.66	60.44
VGG16+Softmax	93.61	91.38	93.23	88.70	66.69
ResNet50+Softmax	92.88	91.45	93.64	90.89	67.15
ResNet101+Softmax	93.95	92.07	94.05	93.51	74.32
BCNN-ResNet50+Softmax	92.73	91.47	93.74	94.27	83.91
BCNN-ResNet101+Softmax	93.90	92.03	93.98	95.18	88.84

BCNN: 双线性卷积神经网络



▲图5 算法识别效果样例

图5是算法识别效果样例,上面两行分别是SYSU-ZTE-CARS数据集上的样例图片和算法预测的属性标签,下面两行分别是在现实应用场景中的测试样例图片和算法预测的属性标签。预测标签中,黑色字体表示正确预测,红色字体表示预测错误。SYSU-ZTE-CARS上正确属性通过构建数据集的时候标注获得,由于现实场景中并没有预标注的属性标签,我们将预测结果与网上搜索结果进行对比来判断预测的正确性。我们抽取了1 000张真实场景的图片作为测试样例,人工评估的结论是:这种跨场景自动标注的方法在非极端条件下的识别精确度较度,比较可靠;在极端条件下,则仍须要人工检验。

4 结束语

基于跨场景标注的思路,通过构建大规模车辆属性数据集,在车辆海报图片场景中训练属性识别模型,再将模型迁移到现实监测控制场景中进行自动标注,以此方式间接获得检测控制场景下的车辆属性标签。通过初步的实验测试评估,该思路可行性较高。

致谢

本研究得到中山大学林惊教授和李冠彬博士的大力支持和帮助,谨

致谢!

实验工作是由中山大学人机物智能融合实验室完成的,周启贤、陈日全同学承担了大量的数据集建立工作和实验,对他们谨致谢意!

参考文献

- [1] 潘秋菱, 许晓杰, 肖乾友. CNGI上的一种电信业务:智能交通[J]. 中兴通讯技术, 2005, 11(5):16-19
- [2] 王添宇, 付超贤. 交通2.0——智慧交通的关键[J]. 中兴通讯技术, 2014, 20(4):11-15. DOI: 10.3969/j.issn.1009-6868.2014.04.003
- [3] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252
- [4] GEIGER A, LENZ P, URSTASUN R. Are We Ready for Autonomous Driving? The Kitti Vision Benchmark Suite[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: IEEE, 2012: 3354-3361
- [5] CAI Z, FAN Q, FERIS R S, et al. A Unified Multi-Scale Deep Convolutional Neural Network for Fast Object Detection[C]// European Conference on Computer Vision. Germany: Springer International Publishing, 2016: 354-370
- [6] REN J, CHEN X, LIU J, et al. Accurate Single Stage Detector Using Recurrent Rolling Convolution[EB/OL].[2017-04-27]. https://arxiv.org/abs/1704.05776
- [7] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[EB/OL].[2017-04-27]. https://arxiv.org/abs/1409.1556
- [8] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2015: 1-9. DOI: 10.1109/CVPR.2015.7298594
- [9] YANG L, LUO P, CHANGELOY C, et al. A Large-Scale Car Dataset for Fine-Grained Categorization and Verification[C]// Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition. USA:IEEE, 2015: 3973-3981. DOI: 10.1109/CVPR.2015.7299023

- [10] SOCHOR J, HEROUT A, HAVEL J. BoxCars: 3D Boxes as CNN Input for Improved Fine-Grained Vehicle Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA:IEEE, 2016: 3006-3015. DOI: 10.1109/CVPR.2016.328
- [11] ZHOU Y, LIU L, SHAO L, et al. DAVE: A Unified Framework for Fast Vehicle Detection and Annotation[C]//European Conference on Computer Vision. Germany: Springer International Publishing, 2016: 278-293
- [12] KRIZHEVSKY A, SUSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks[C]// Advances in Neural Information Processing Systems. USA: IEEE, 2012: 1097-1105
- [13] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2016: 770-778. DOI: 10.1109/CVPR.2016.90
- [14] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN Models for Fine-Grained Visual Recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. USA:IEEE, 2015: 1449-1457
- [15] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. USA:ACM, 2014: 675-678

作者简介

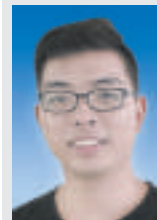


董振江, 中兴通讯股份有限公司战略与技术专家委员会业务专家组组长、云计算及IT研究院副院长, 中国人工智能学会常务理事; 主要研究方向为云计算与大数据、新媒体、移动互联网等; 曾获国家科技进步二等奖2次、省部级奖项多次, 主持基金项目10

余项; 已发表论文20余篇, 出版专著1部。



高燕, 中兴通讯有限公司高级工程师; 主要研究领域包括人工智能、图像和视频分析; 主持5项中兴通讯产学研合作项目, 参与3项国家重大科技项目; 发表论文多篇。



吴文熙, 中山大学数据科学与计算机学院研究生; 主要研究领域为深度学习与计算机视觉, 主要研究课题为大规模物体识别; 在Kaggle、ImageNet等国际计算机视觉比赛中均有优秀表现。

领域自适应目标识别综述

Domain Adaptation Object Recognition

唐宋/TANG Song
叶茂/YE Mao
李旭冬/LI Xudong

(电子科技大学, 四川 成都 611731)
(University of Electronic Science and
Technology of China, Chengdu 611731,
China)

在计算机视觉领域,作为许多计算机视觉应用的基础,目标识别问题一直是相关研究的重点。2006年, G.E.Hiton 提出深度学习以后,针对经典的目标识别问题(训练集和测试集概率分布相同),取得了重大进展。基于域间分布差异的域自适应目标识别问题,已逐渐成为计算机视觉领域的研究热点。对于该问题的研究具有重要的意义,主要体现在两个方面:提高了分类器或检测器的复用性,有效地增强了模型对新环境的适应性;使得模型的训练过程能很大程度地独立于应用场景。因为在训练模型时,不再特别地考虑应用场景的具体特点,所以在训练阶段可以借助诸如大数据等的优势,预训练出具有较好泛化能力的模型。

1 域自适应学习和目标识别问题

域自适应学习是迁移学习^[1]的一

收稿日期: 2017-05-28

网络出版日期: 2017-07-17

基金项目: 国家自然科学基金
(61375038); 四川省科技厅基础应用研究
计划(2016JY0088)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0025-007

摘要: 针对目前域自适应目标识别问题的学习方法,进行系统总结。首先,提出目标识别的两个基本主题:基于域自适应的目标分类和目标检测;然后,围绕这两个主题,从特征和样本两个角度,展开具体综述。认为对于域自适应目标分类,几种算法的主要问题为:忽略了样本所构成的流形几何结构,如果能利用几何结构来约束特征表达,将有利于样本特征鲁棒性的提高。对于域自适应目标检测,其问题为:现有方法对源样本和带标签的目标域样本存在依赖,这一问题使得现有的方法很难适用于某些真实的应用场景。

关键词: 域自适应学习;目标分类;目标检测

Abstract: In this paper, a systematic review on domain adaptation of object detection is presented. At first, the two basic subtopics of object recognition—object classification and object detection based on domain adaption are proposed; and then, from the views of feature and samples, two problems are reviewed in detail. For the domain adaptation object classification, the existing methods ignore the manifold structure of samples. In fact, the geometric information is helpful to obtain robust representation. For the domain adaptation object detection, the existing methods depend on the source samples and labeled target samples, which makes these methods hard to be employed into some real applications.

Keywords: domain adaptation study; object classification; object detection

个子类。对于域自适应学习,源域和目标域的任务相同,但是,数据不同但相关。这类学习的核心任务是解决两个域数据分布的差异问题,是迁移学习最基本的形式。

目标识别问题包含狭义和广义两个层面:狭义是指目标分类,其主要任务是将不同类别的物体彼此区分开;广义是指目标检测,要明确图片中存不存在目标物体?这些目标物体的具体位置在哪里?目前,常用的目标检测思路大概大概有3种,它们分别为:

(1)滑动窗口机制+分类器的检测框架^[2-3],检测过程如图1所示。因为具有极强的通用性,早期的检测方

法多采用该框架,目前该框架仍然发挥着重要的作用。但是,由滑动窗口机制产生的候选检测窗有时很难准确地框定目标,容易造成漏报情况的发生。

(2)卷积神经网络(CNN)的回归检测框架^[4-7],如图2所示。该框架将测试图片作为整体进行认知,不仅有效解决了第1种框架目标丢失的问题,而且能方便地引入了上下文信息,取得了较好的检测结果。但是,因该框架只适用于CNN,所以其通用性稍差,并且相对于其他方法,样本标记相对麻烦。

(3)基于区域分割的检测框架。该框架由文献[8-10]于2014年提出,

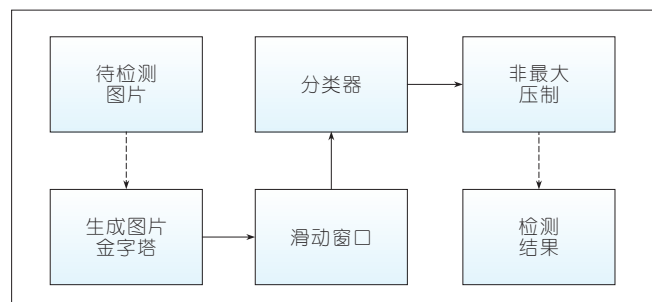


图1
滑动窗口机制+分类器的
检测框架检测流程
示意

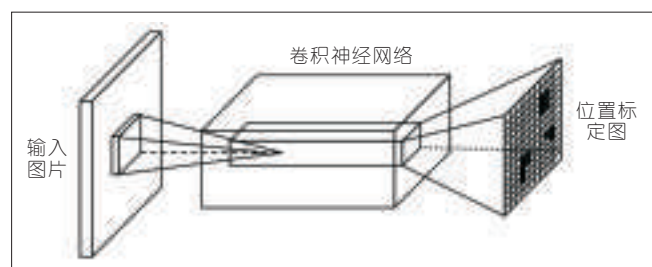


图2
基于CNN的回归检测
框架结构示意

大致检测流程如图3所示。首先,对输入图片采用选择性搜索方法提取出区域候选框;然后,所有候选框缩放成固定大小;接着,利用CNN提取特征;最后,用支持向量机(SVM)进行区域分类,并采用回归的方法微调候选框位置与大小。该框架主要的贡献是去掉了窗口扫描,用聚类方式,对图像进行分割分组,得到多个候选框的层次组。

通过以上介绍,可得出两点结论:目标分类是目标检测的基础和前提;目标检测是目标分类的扩展,在目标分类的基础上引入不同的目标定位方法就形成了不同的检测思路。因此,针对目标识别的相关研究需要兼顾目标分类和目标检测这两个层次。

基于上述分析,我们将域自适应目标识别问题细化为两个基本研究主题:基于域自适应的目标分类和目标检测。为了聚焦讨论重点,在文章中,我们仅考虑只有一个源域的情形,而对多源域的情形则不做详细的讨论。

2 域自适应目标分类方法

现有域自适应目标分类方法可以分为:半监督域自适应方法和无监

督域自适应方法。

2.1 半监督领域自适应方法

在这种场景中,源域中的样本都含有标签信息,目标域中的样本只有一部分含有标签。这类方法的核心思想是利用源域和目标域的标签信息,构建源域和目标域之间的关系。这些方法可分为基于特征和基于样本两类。基于特征的半监督方法,比较代表性的论文有文献[11]和[12]:文献[11]提出了一种度量学习算法,该算法利用源域和目标域中所有带标签的样本,学习一个线性变换,在该变换的作用下,所有样本被映射到一个具有域不变特性的特征空间中;文献[12]给出了一个字典学习算法,基于两个域中带标签的样本,由该字典编码的样本特征同时具有域不变性和稀疏性。

基于样本的半监督方法,比较代

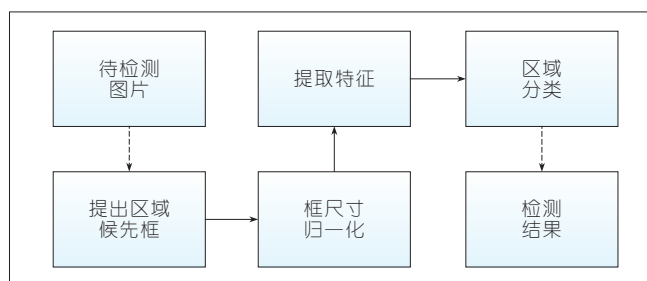
表性的论文有文献[13-17]:文献[13]中提出利用标记的目标样本,来确定被弱标记的源域样本的正确标签,然后基于这些有标签的源域和目标域样本,训练分类器;文献[14]提出在传统的 Least-Squares SVM 中引入了两个新的正则化项,一个利用了连续性假设,另一个由带标记的目标样本构成,以此来描述目标分类器的经验误差;文献[15-17]提出了基于 Adaboost 的学习算法,在算法实现中,为了判别源样本是否对目标域学习有用,必须借助少量带标签的目标域样本,来训练适用于目标域的分类器,并将其作为判别器;文献[18]总结了之前基于 SVM 实现知识迁移的方法,利用源领域获得支持向量迁移知识,并在此基础上,设计了称为 Cross domain SVM 的支持向量机模型,该方法需要目标领域的部分样本带有类别标签。

2.2 无监督领域自适应方法

在这类方法中,只有源域中的样本含有标签信息,而目标域中的样本没有标签。与上述半监督方法相比,无监督方法难度更大。与半监督方法相同,这些方法也可以归纳为基于特征和基于样本两类,为了叙述方便,将它们分别记为 A 类和 B 类。具体地, A 类算法假设源域和目标域的样本能被映射到共享的隐藏特征空间,在这个空间中,源域样本和目标域样本的特征概率分布差异很小。由这一思想衍生出两种不同解决方案,分别记为 A1 和 A2。

A1 方案从直接刻画该隐藏特征空间的角度切入问题(原理如图4所示),假设源域和目标域对应的特征

图3
基于语义分割的检测
框架的流程示意



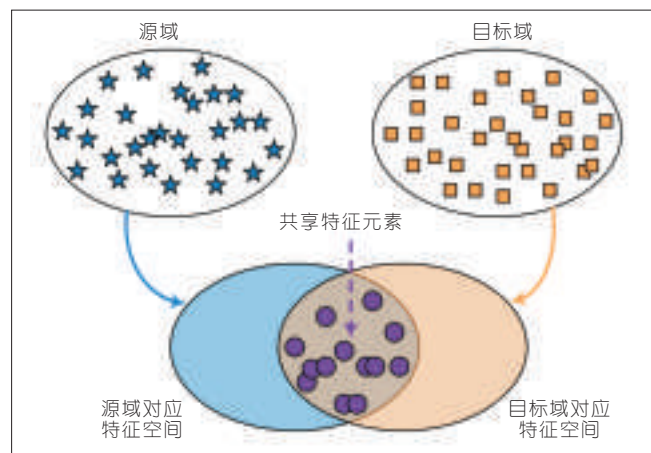


图4
无监督领域自适应方法中基于特征算法方案1(A1)的原理示意

空间为同构空间,而且在特征层面两者存在一些共享的特征元素;该隐藏特征空间可以通过这些共享特征元素加以描述。对于这一思路,其核心是如何找到刻画该隐藏特征空间的共享特征元素。比较代表性的论文有文献[19-23]:文献[19]提出了基于中枢特征的方法,通过提取最重要的中枢特征,形成辅助特征向量,最后分类器在原特征向量加上该辅助特征向量的空间内重新学习,该方法的问题在于当目标领域未知时,无法选择合适的中枢特征;基于类似思想,文献[20-22]分别基于互聚类方法、降维方法以及多核学习的方法实现了分类器的迁移;文献[23]利用跨域边界样本选择过程,构建域不变特征子空间。

A2方案则聚焦到源域和目标域对应的特征空间为异构空间时的情形,其基本思想是通过学习一个变换关系将源领域与目标领域特征空间映射到该隐藏特征空间(原理如图5所示)。比较代表性的论文有文献[24-29]:文献[24]中提出利用多个视角的数据构建翻译器,通过这个翻译器,实现源领域与目标领域特征空间的映射;文献[25]中提出通过非线性度量学习方法得到一个非对称的特征空间映射;文献[26]中提出将源领域与目标领域样本映射到同一子空间,通过SVM学习,训练得到新的目标分类器和特征空间的映射;文献

[27]中提出基于最大平均差异(MMD)概念,通过对源域和目标域上的边缘分布和条件分布进行约束,来实现对变换矩阵的求解;文献[28-29]中提出了基于深度模型的方法,其思想是借助深度模型强有力的特征抽象能力,通过多层网络结构,在不断降维抽象的同时,不断地弥合源域和目标域概率分布的差异,在最高特征抽象层使两个域之间的概率分布差异最小。

文献[30-31]中提出了一类基于MMD的深度网络方法。具体地,文献[30]中设计的深度自适应网络(DAN)由两部分构成:首先,利用深度卷积网络提取共享特征;然后,构建两个全链接子网络,来分别提取源域和目标域的域特征,同时,在这两

个子网络中,逐层引入针对多核可变的最大平均差异(MK-MMD)最小化约束,从而实现向隐藏特征空间的映射。基于类似的思想,文献[31]中将MMD、残差网络结合,使用熵最小化方法,提出了自适应残差迁移网络。

最近,文献[32]提出了一种更为自然的迁移方案:结合生成对抗网络框架^[33]与深度学习来解决领域自适应(DA)问题,所提深度网络能够学到具有领域不变性的特征。文献[34-35]则从传统的核映射角度提出了基于特征的迁移方法,其基本思想是用高维特征空间来表示该隐藏特征空间,而源样本和目标域样本在该高维空间的投影可以用核映射方法得到,基于这些投影点可以建立相应的学习模型。

B类方法将样本本身的特点作为解决问题的切入点,大致可以分为两类,分别记为B1和B2。具体地,B1类方法的基本思想如图6所示(假设源域中的一部分样本满足目标域概率分布),这类方法通常采用加权的方式,来筛选出满足条件的源样本,以此训练出适用于目标域的分类器,因此其核心问题是如何求取这些加权系数。比较代表性的论文有文献[36-38]:文献[36]中提出了一个基于样本加权的方法,使源域样本的概率分布最大可能地匹配目标域的概率

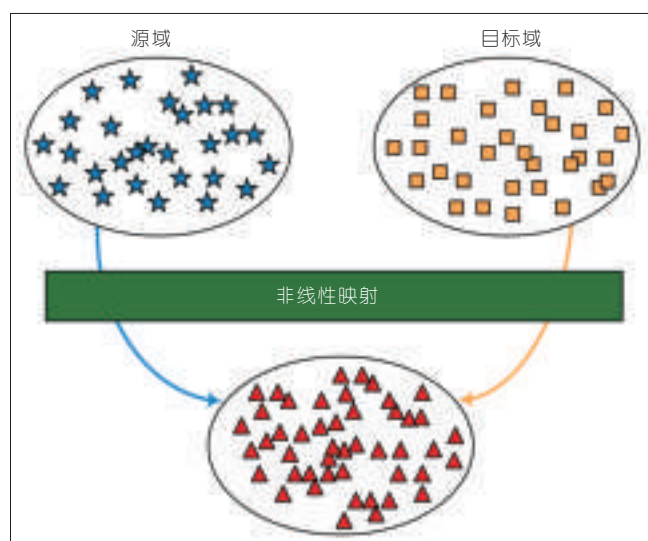


图5
无监督领域自适应方法中基于特征算法方案2(A2)的原理示意

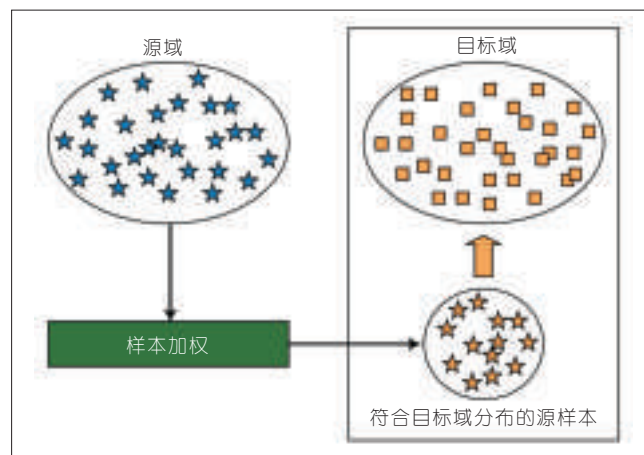


图6
无监督领域自适应
方法中基于样本算法
方案1(B1)的原理示意

分布,然后利用加权后的源域样本和目标域样本完成迁移;文献[37]则提出直接利用平均相似度的方法,直接将符合目标域概率分布的源域样本筛选出来;文献[38-39]估计源领域与目标领域样本密度函数,然后通过计算两者之比,来获得权重。为避免估计密度函数,文献[40]提出采用K-L散度估计该比值。

B2类方法的基本思想是:如果能有效刻画源域到目标域的迁移过程,那么源样本空间则能以渐进的方式,投影到目标域空间,其原理如图7所示。比较代表性的论文有文献[41-46]:文献[41]中提出利用流形对齐过程,学习从源样本空间到目标样本空间的变换,实现源空间到目标空间的投影;文献[42]中提出假设源域样本和目标域样本所构成的空间是位于Grassman流形上的两个不同的点,在这个Grassman流形上,连接这两个点的测地线描述了源域到目标域的迁移过程,通过子空间插值可拟合出这条测地线,将样本投影到这些子空间上,由所有这些投影所构成的向量被作为样本的特征,该文献为域自适应的问题提供了一个新的观察角度;通过对测地线上子空间的积分,文献[43]中提出了被称为GFK的核方法;文献[44]中提出可以按比例构成测地线上子空间的样本集,并利用这些样本集训练得到基于深度模型分类器,相比之前的线性投影,样本投影到每

个子空间上的特征,具有更强的表达能力;文献[45]中则提出通过增量学习的方式,为测地线上每个子空间建立相应的字典,然后利用这些字典完成对样本的编码;文献[46]则提出在Spline流形上对测地线进行子空间的采样。

3 域自适应目标检测方法

目前,域自适应学习在计算机视觉中的应用中已有一些研究成果^[47],但是关于域自适应的目标检测方法研究并不是很多。现有的工作大致可以分为两类:半监督域自适应目标检测方法和无监督域自适应目标检测方法。为方便叙述,它们被分别记为C和D。

3.1 半监督域自适应目标检测方法

在域自适应目标检测问题中,半监督是指:模型训练过程能够从目标应用场景中获得一些带标签信息的

样本。这些方法几乎都是基于特征的方法,其基本思想是通过带标签的目标域样本,抽取跨域的特征(基本框架如图8a)所示)。根据源领域是否额外引入大规模数据集,这类方法可被划分为C1和C2两个子类。

C1类方法是相对传统的方案,在相应的问题假设中,它们使用的源域数据集通常规模较小。比较有代表性的论文有文献[48-50]:文献[48]中提出首先利用卷积自编码器在源域上预训练卷积核,然后基于带标签的目标域样本,对上述卷积核进行调谐(FT),使之适用于目标域;文献[49]提出面向一种车辆检测的CNN方法,该方法保留源域和目标域共享的卷积核的同时,借助带标签的目标域样本对非共享卷积核进行更新;文献[50]中提出通过对源领域特征迁移构造弱分类器,并将每个弱分类器已学习好的权重根据目标场景再学习调整,通过这种方式,能解决由于视角以及光照等造成的改变。

C2方法是最近才逐渐发展起来。这类方法利用在大数据上训练好的深度模型,来提取源样本的深度特征,然后基于所得深度特征集合进行算法设计。相比于传统的C1类方法,这类方法检测效果更好,主要有两方面原因:一方面,有效融合了深度特征良好的区分能力和迁移能力;另一方面,可以方便地和其他算法框架相结合。比较有代表性的论文有文献[8-10]、文献[52-53]:文献[8-10]提出并发展了一类基于区域分割的

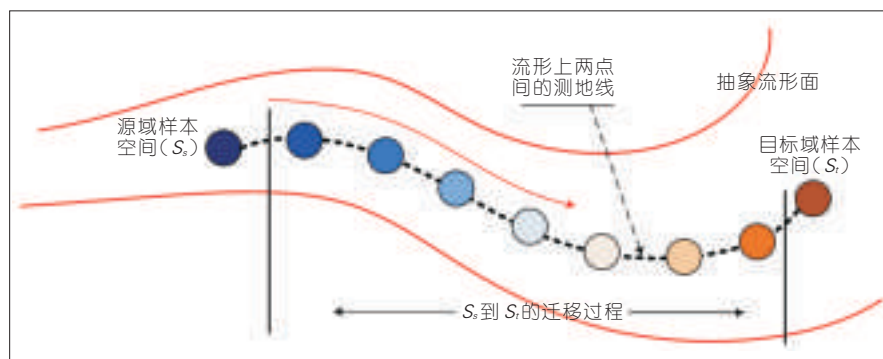
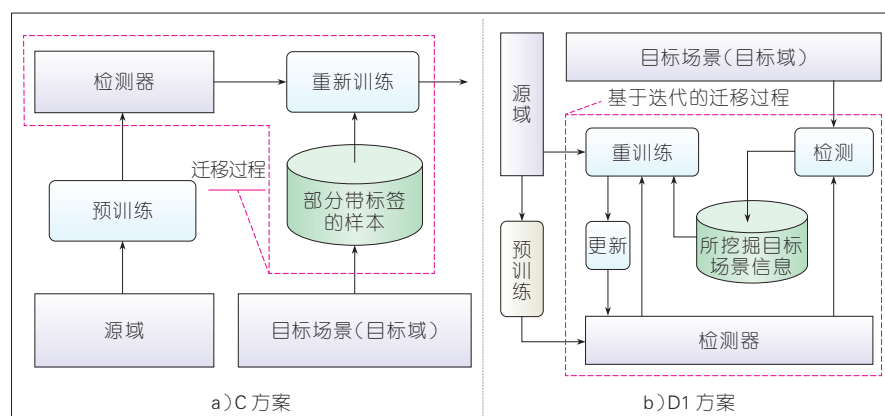


图7 无监督领域自适应方法中基于样本算法方案2(B2)的原理示意



▲图8 域自适应目标检测方法原理示意

多类目标检测方法,这类方法通过已训练好的深度模型来取得特征^[51],同时结合SVM方法来实现类别判定;文献[52-53]中提出利用已训练好的深度模型对整个检测场景提取深度特征,并结合长短期记忆网络(LSTM)模型,提出基于序列分割的回归检测方法。

3.2 无监督域自适应目标检测方法

无监督的含义是指目标应用场景不提供任何带标签的样本。这些方法几乎都是基于实例的,也是目前学界使用较多的路线,它们大致可以分为两类,分别记为D1和D2。

D1类方法基本思路是通过迭代的方式,从目标域中挖掘有用信息,来重新训练原检测器,其基本框架如图8b)所示。比较有代表性的论文有文献[54-61]:文献[54]中提出通过一个基于外表特征的目标检测器,来设定置信度,将目标场景中满足置信度的正负样本挑选出来,重新训练检测器;文献[55-56]利用互训练方法,迭代地训练基于不同特征的分类器;文献[57]提出一种在线分类器学习算法,在每次迭代中,该算法通过背景差化技术,对目标场景检测结果进行自动标注,然后借助这些标注结果重新训练原分类器;文献[58]结合复杂网络,迭代学习目标检测器以及源领域和目标领域样本的权重;文献[59-60]中基于图的方法、上下文信息计

算源领域和目标领域样本的置信度,通过融合置信度和上下文信息的SVM,不断迭代训练样本权重和置信度,取得较好的效果;与之类似,文献[61]提出了一个基于深度CNN的方法,通过新增重建层和聚类层,实现了对多尺度场景特征和视觉模式的挖掘。

D2类方法是非迭代的方案。比较有代表性的论文有文献[62-63]:文献[62]提出了基于特征联合学习的深度网络,该方法中首先基于源域样本和目标域样本,通过栈式自编码器(SAE),学习得到一个特征变换,将样本映射到某特征空间,然后利用源域样本的标签信息,训练得到基于该特征空间的SVM分类器;文献[63]提出了针对监测控制视频的检测器迁移方法,它的核心思想是通过视频的时序线索来对随机生成的候选图像块进行筛选,然后利用过滤后的样本训练新的分类器;文献[64]提出一种基于神经调控的检测器迁移方法,该方法利用一个调控网络对位于深度网络末端的分类器,进行自适应地加权调整,为每一个目标域样本,生成专有的分类器。

4 问题和展望

对于域自适应目标分类,主要问题为:现有的方法要么基于概率匹配的思路(A类和B1类),要么从抽象流形这一个几何观点设计相关算法

(B2类),它们都忽略了样本本身所构成的流形几何结构。事实上,样本本身是存在某种几何关系的,即所谓的流形结构。大量与流形学习有关的工作^[65-66]已经证实了这一点。如果我们能利用这些几何结构来约束特征表达,将有利于样本特征鲁棒性的提高。

对于域自适应目标检测,其问题为:现有方法为了迁移检测器,要么要求目标场景含有带标签的样本,来实现对源域知识的迁移(C类);要么需要保留全部源样本,来实现对目标场景挖掘信息的评估(D类)。换言之,现有方法对源样本和带标签的目标域样本存在依赖。这一问题使得现有的方法很难适用于某些真实的应用场景,这主要体现在3个方面:(1)有些真实场景很难获取带标签的样本,特别是非结构化的应用环境;(2)大量硬件设备的计算能力和存储空间都是有限的,基于所有源样本进行置信度评估不仅会造成额外的计算开销,还会带来巨大的存储负担;(3)对具有动态背景的场景并不适用。

5 结束语

近年来,许多学者针对域自适应目标识别问题,从理论和应用方面展开研究,并取得很多成果。文章对现有的域自适应目标识别方法进行了综述:首先,我们分别介绍域自适应学习和目标识别,在此基础上,将域自适应目标识别细化为域自适应分类和域自适应目标检测两个基本问题;然后,以半监督和无监督为基本线索,分别对这两个问题进行了综述,重点对各种方法的核心思想,以及发展脉络进行了梳理;最后,总结了现有方法所存在的问题。

参考文献

- [1] PAN J S, YANG Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering (TKDE), 2010, 22(10): 1345-1359. DOI: 10.1109/TKDE.2009.191

- [2] HOWSE J. OpenCV Computer Vision with Python[M]. UK: Packt Publishing Ltd., 2013
- [3] NEUBECK A, GOOL L V. Efficient Non-Maximum Suppression[C]//International Conference on Pattern Recognition (ICPR). China, 2006: 850-855
- [4] SZEGEDY C, TOSHEV A, ERHAN D. Deep Neural Networks for Object Detection[C]//Annual Conference on Neural Information Processing Systems (NIPS). USA, 2013: 2553-2561
- [5] GIRSHICK R, IANDOLA F, DARRELL T, et al. Deformable Part Models are Convolutional Neural Networks[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA:IEEE, 2015: 437-446
- [6] REDOM J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA:IEEE, 2016:779-788
- [7] FU M, XU P, LI X, et al. Fast Crowd Density Estimation with Convolutional Neural Networks[J].Engineering Applications of Artificial Intelligence, 2015, 43:81-88. DOI: 10.1016/j.engappai.2015.04.006
- [8] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). USA:IEEE, 2014:580-587
- [9] GIRSHICK R. Fast R-CNN[C]//IEEE International Conference on Computer Vision (ICCV). USA: IEEE, 2015: 1440-1448
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[C]//Annual Conference on Neural Information Processing Systems (NIPS). Canada: NIPS, 2015: 91-99
- [11] KATE S, BRIAN K, MARIO F, et al. Adapting Visual Category Models to New Domains [C]//European Conference on Computer Vision (ECCV). Greece: ECCV, 2010:213-226
- [12] QIU Q, PATEL V M, TURAGA P, et al. Domain Adaptive Dictionary Learning [C]//European Conference on Computer Vision (ECCV). Italy: ECCV, 2012, 631-645
- [13] BERGAMO A, TORRESANI L. Exploiting Weakly-Labeled Web Images to Improve Object Classification: A Domain Adaptation Approach [J]. Annual Conference on Neural Information Processing Systems (NIPS), 2010:181-189
- [14] DUAN L, TSANG I W, XU D, et al. Domain Adaptation from Multiple Sources Via Auxiliary Classifiers[C]//International Conference on Machine Learning (ICML). Canada: ICML, 2009:289-296
- [15] DAI W, YANG Q, XUE G R, et al. Boosting for Transfer Learning[C]//International Conference on Machine Learning (ICML). USA: ICML, 2007: 193-200
- [16] XU Z, SUN Z. Multi-View Transfer Learning with Adaboost [C]//International Conference on Tools with Artificial Intelligence (ICTAI). USA: ICTAI, 2011:399-402
- [17] XU Z, SUN S. Multi-Source Transfer Learning with Multi-View Adaboost [C]//International Conference on Neural Information Processing (ICNIP). Qatar: ICNIP, 2012:332-339. DOI: 10.1007/978-3-642-34487-9_41
- [18] JIANG W, ZAVESKY E, CHANG S F, et al. Cross-Domain Learning Methods for High-Level Visual Concept Classification[C]//IEEE International Conference on Image Processing (ICIP). USA:IEEE, 2008:161-164
- [19] BLITER J, McDONALD R, PEREIRA F. Domain Adaptation with Structural Correspondence Learning[C]//Conference on Empirical Methods in Natural Language Processing (EMNLP). Australia: EMNLP, 2006:120-128
- [20] DAI W, XUE G R, YANG Q, et al. Co-Clustering Based Classification for Out-of-Domain Documents[C]//International Conference on Knowledge Discovery and Data Mining (ICKDDM), USA: ICKDDM, 2007:210-219
- [21] PAN S J, I. TSANG W, KWOK J T, et al. Domain Adaptation Via Transfer Component Analysis[J].IEEE Transactions on Neural Networks (TNN), 2011, 22(2):199-210
- [22] DUAN L, TSANG I W, XU D. Domain Transfer Multiple Kernel Learning [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI), 2011, 34(99):1-1
- [23] HUBERT T Y H, YEH Y R, WANG F Y C. Learning Cross-Domain Landmarks for Heterogeneous Domain Adaptation[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA:IEEE, 2016:5081-5090
- [24] DAI W, CHEN Y, XUE G R, et al. Translated Learning: Transfer Learning Across Different Feature Spaces[C]//Annual Conference on Neural Information Processing Systems (NIPS). Canada: NIPS, 2008:353-360
- [25] KULIS B, SAENKO K, DARRELL T. What You Saw is Not What You Get: Domain Adaptation Using Asymmetric Kernel Transforms[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR).USA: CVPR, 2011:1785-1792. DOI: 10.1109/CVPR.2011.5995702
- [26] DUAN L, XU D, TSANG I. Learning with Augmented Features for Heterogeneous Domain Adaptation [J]. Computer Science, 2012:711-718. DOI: 10.1109/TPAMI.2013.167
- [27] LONG M, WANG J, DING G, et al. Transfer Feature Learning with Joint Distribution Adaptation[C]//IEEE International Conference on Computer Vision (ICCV). Australia: ICCV, 2013: 2200-2207. DOI: 10.1109/ICCV.2013.274
- [28] M. Chen, Z. Xu, K. Weinberger, et al. Marginalized Denoising Autoencoders for Domain Adaptation[J]. Computer Science, 2012
- [29] GLOROT X, BORDES A, BENGIO Y. Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach[C]//International Conference of Machine Learning (ICML). USA:ICML, 2011: 611-618
- [30] LONG M J W, CAO Y, JORDAN M I. Learning Transferable Features with Deep Adaptation Networks[C]//International Conference on Machine Learning (ICML), France: ICML, 2015: 97-105
- [31] LONG J W M, ZHUY H, JORDAN M I. Unsupervised Domain Adaptation with Residual Transfer Networks[C]//Annual Conference on Neural Information Processing Systems (NIPS). Spain: NIPS, 2016:10-18
- [32] GANIN V L Y. Unsupervised Domain Adaptation by Back Propagation[C]//International Conference on Machine Learning (ICML). France: ICML, 2015:280-288
- [33] GOODFELLOW I J, POUGETABADIE J, MIRZA M, et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014(3): 2672-2680
- [34] SUN S, XU Z, YANG M. Transfer Learning with Part-Based Ensembles[C]//International Workshop on Multiple Classifier Systems. China: LNCS, 2013:271-282. DOI: 10.1007/978-3-642-38067-9_24
- [35] TU W, SUN S. Crossdomain Representation-Learning Framework with Combination of Class Separate and Domain Merge Objectives[C]//International Workshop on Cross Domain Knowledge Discovery in Web and Social Network Mining. USA, 2012:18-25. DOI: 10.1145/2351333.2351336
- [36] SCHOLKOPF B, PLATT J, HOFMANN T. Correcting Sample Selection Bias by Unlabeled Data[C]//Annual Conference on Neural Information Processing Systems (NIPS). Canada: NIPS, 2006: 601-608
- [37] GONG B, GRAUMAN K, SHA F. Connecting the Dots with Landmarks: Discriminatively Learning Domain-Invariant Features for Unsupervised Domain Adaptation[C]//International Conference on Machine Learning (ICML). USA:ICML, 2013:222-230
- [38] ZADROZNY B. Learning and Evaluating Classifiers Under Sample Selection Bias[C]//International Conference on Machine Learning (ICML).Canada: ICML, 2004: 114-122
- [39] FAN W, DAVIDSON I, ZADROZNY B, et al. An Improved Categorization of Classifier's Sensitivity on Sample Selection Bias[C]//International Conference on Data Mining. USA, 2005:4-10
- [40] SUGIYAMA M, NAKAJIMA S, KASHIMA H, et al. Direct Importance Estimation with Model Selection and Its Application to Covariate Shift Adaptation[C]//Annual Conference on Neural Information Processing Systems (NIPS). Canada: NIPS, 2008:1433-1440
- [41] WANG C, MAHADEVAN S. Heterogeneous Domain Adaptation Using Manifold Alignment[C]//International Joint Conference on Artificial Intelligence (IJCAI), Spain: IJCAI, 2011:1541
- [42] GOPALAN R, LI R, R. Chellappa. Domain Adaptation for Object Recognition: An Unsupervised Approach[C]//IEEE International Conference on Computer Vision (ICCV), Spain, 2011: 999-1006. DOI: 10.1109/ICCV.2011.6126344
- [43] SHA F, SHI Y, GONG B, et al. Geodesic Flow Kernel for Unsupervised Domain Adaptation[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA:IEEE, 2012:2066-2073
- [44] CHOPRA R G S, BALAKRISHNAN S. DLID:

- Deep Learning for Domain Adaptation by Interpolating Between Domains[C]//International Conference on Machine Learning (ICML) Workshop on Challenges in Representation Learning. USA: ICML, 2013
- [45] NI J, QIU Q, CHELLAPPA R. Subspace Interpolation Via Dictionary Learning for Unsupervised Domain Adaptation[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA:IEEE, 2013: 692–699
- [46] CASEIRO R, HENRIQUES J F, MARTINS P, et al. Beyond the Shortest Path: Unsupervised Domain Adaptation by Sampling Subspaces Along the Spline Flow [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA:IEEE, 2015:3846–3854. DOI: 10.1109/CVPR.2015.7299009
- [47] DUAN L, XU D, TSANG W H, et al. Visual Event Recognition in Videos by Learning from Web Data[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: IEEE, 2010: 1959–1966. DOI: 10.1109/TPAMI.2011.265
- [48] SERMANET P, KAVUKCUOGLU K, CHINTALA S, et al. Pedestrian Detection with Unsupervised Multistage Feature Learning[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: IEEE, 2013:3626–3633
- [49] LI X, YE M, FU M, et al. Domain Adaption of Vehicle Detector Based on Convolutional Neural Networks [J]//International Journal of Control, Automation and Systems (IJCAS), 2015, 13(4):1020–1031. DOI: 10.1007/s12555-014-0119-z
- [50] PANG J, HUANG Q, YAN S, et al. Transferring Boosted Detectors Towards Viewpoint and Scene Adaptiveness [J]. IEEE Transactions on Image Processing (TIP), 2011, 20(5):1388–1400
- [51] DENG J, DONG W, SOCHER R, et al. ImageNet: A Large-Scale Hierarchical Image Database[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: IEEE, 2009: 248–255. DOI: 10.1109/CVPR.2009.5206848
- [52] LI X, YE M, LIU D, et al. Memory-Based Object Detection in Surveillance Scenes[C]//IEEE International Conference on Multimedia and Expo (ICME). USA,IEEE, 2016:1–6. DOI: 10.1109/ICME.2016.7552920
- [53] LI X, YE M, LIU Y, et al. Accurate Object Detection Using Memory-Based Models in Surveillance Scenes[J]. Pattern Recognition, 2017(67):73–84
- [54] ROSENBERG C, HEBERT M, SCHNEIDERMAN H. Semi-Supervised Self-Training of Object Detection Models [C]//IEEE Workshop on Applications of Computer Vision/IEEE Workshop on Motion and Video Computing. USA:IEEE, 2005:29–36
- [55] LEVIN A, VIOLA P, FREUND Y. Unsupervised Improvement of Visual Detectors Using Co-Training[C]//IEEE International Conference on Computer Vision (ICCV). USA: IEEE, 2003:626–633
- [56] WU B, NEVATIA R. Improving Part Based Object Detection by Unsupervised, Online Boosting[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: IEEE, 2007:1–8
- [57] NAIR V, CLARK J. An Unsupervised, Online Learning Framework for Moving Object Detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA:IEEE, 2004: 317–324. DOI: 10.1109/CVPR.2004.34
- [58] WANG M, WANG X. Automatic Adaptation of a Generic Pedestrian Detector to a Specific Traffic Scene[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA:IEEE, 2011: 3401–3408. DOI: 10.1109/CVPR.2011.5995698
- [59] WANG X, LI W, WANG M. Transferring a Generic Pedestrian Detector Towards Specific Scenes[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA:IEEE, 2012:3274–3281
- [60] WANG X, WANG M, LI W. Scene-Specific Pedestrian Detection for Static Video Surveillance[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI), 2014, 36(2):361–74
- [61] ZENG X, OUYANG W, WANG M, et al. Deep Learning of Scene-Specific Classifier for Pedestrian Detection[C]//European Conference on Computer Vision (ECCV). Switzerland:ECCV, 2014:472–487
- [62] HTIKE K K, HOGG D. Unsupervised Detector Adaptation by Joint Dataset Feature Learning[C]//International Conference on Computer Vision and Graphics (ICCVG). Poland: ICCVG, 2014: 270–277
- [63] HTIKE K K, Hogg D C. Efficient Non-Iterative Domain Adaptation of Pedestrian Detectors to Video Scenes[C]//IEEE International Conference on Pattern Recognition (ICPR). USA:IEEE, 2014:654–659. DOI: 10.1109/ICPR.2014.123
- [64] TANG S, YE M, ZHU C, et al. Adaptive Pedestrian Detection Using Convolutional Neural Network with Dynamically Adjusted Classifier[J]. Journal of Electronic Imaging, 2017, 26(1): 013012
- [65] ROWEIS S T, SAUL L K. Nonlinear Dimensionality Reduction by Locally Linear Embedding [J]. Science, 2000, 290(5500): 2323–2326
- [66] TENENBAUM J B, SILVA V D, LANGFORD J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. Science, 2001, 290(5500):2319–2323

作者简介



唐宋, 电子科技大学博士; 主要研究领域为深度学习、模式识别、非线性系统控制; 先后参加基金项目5项; 已发表论文20余篇, 其中被SCI/EI检索10余篇。



叶茂, 电子科技大学教授、博士生导师; 主要研究领域为机器学习与计算机视觉; 先后主持和参加基金项目10余项; 已发表论文80余篇, 其中被SCI检索50余篇。



李旭冬, 电子科技大学博士; 主要研究领域为深度学习、模式识别; 先后参加基金项目2项; 已发表论文10余篇, 其中被SCI/EI检索5篇。

深度神经网络学习的结构基础: 自动编码器与限制玻尔兹曼机

Architectures of Deep Neural Networks: Auto-Encoders and Restricted Boltzmann Machines

康文斌 / KANG Wenbin¹

彭菁 / PENG Jing²

唐乾元 / TANG Qianyuan³

(1. 湖北医药学院, 湖北 十堰 442000;
2. 平安科技(深圳)有限公司, 广东 深圳 518000;

3. 香港浸会大学, 香港 九龙塘 999077)
(1. Hubei University of Medicine, Shiyan 442000, China;

2. Ping An Technology (Shenzhen) Co., Ltd., Shenzhen 518000, China;

3. Hong Kong Baptist University, Hong Kong 999077, China)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0032-004

摘要: 自动编码器(AE)和限制玻尔兹曼机(RBM)是在深度神经网络领域广泛使用的两种常见的基础性结构。它们都可以作为无监督学习的框架,通过最小化重构误差,提取系统的重要特征;更重要的是,通过多层的堆叠和逐层的预训练,层叠式自动编码器和深度信念网络都可以在后续监督学习的过程中,帮助整个神经网络更好地收敛到最小值点。

关键词: 深度学习; 神经网络; AE; RBM

Abstract: Auto-encoders (AE) and Restricted Boltzmann Machines (RBM) are two kinds of basic building blocks which are widely used in the architectures of deep neural networks. By minimizing the reconstruction errors, both the AE and the RBM can extract the key characteristics of the input data and can work as the basic framework of the unsupervised learning. Moreover, with the layer-by-layer stacking and layer-wise pre-training, both the stacked AE and the deep belief networks can help neural networks converge faster and better in the following supervised fine-tuning process.

Key words: deep learning; neural network; AE; RBM

近年来,深度学习在图像和语音识别、自然语言处理、推荐系统等诸多领域中取得了许多重要的突破,深度学习的许多重大进展为解决许多长期以来难以解决的困难问题提供了崭新的思路^[1-3]。深度学习以人工神经网络为结构基础,在一个神经网络中,如图 1a)所示,每个神经元都是一个感知机,输入端的数据在线性组合后,经过激活函数引入了非线性因素。在一个神经网络的输入层和输出层之间常常会有一个或者多个隐藏层,如图 1b)和 c)中所示。通过许多个包含不同连接权重的感知机的组合和叠加,一个神经网络因而具有了极强的表示能力。“深度学习”这一名词中的深度指的是神经网络

中隐藏层的数量。多个隐藏层让深度神经网络能够表示数据中更为复杂的特征,例如:在用深度卷积神经网络(CNN)进行人脸识别时,较为底层的隐藏层首先提取的是图片中一些边缘和界面的特征,随着层级的提高,图片中一些纹理的特征可能会显现,而随着层级继续提高,一些具体的对象将会显现,例如:眼睛、鼻子、耳朵等,再到更高层时,整个人脸的特征也就被提取了出来。在一个深度神经网络上,较高层的特征是低层特征的组合,而随着神经网络从低层到高层,其提取的特征也越来越抽象、越来越涉及“整体”的性质^[4]。

神经网络的训练在本质上是一个非线性优化问题,要求在已知的约束条件下,寻找一组参数组合,使该组合确定的目标函数达到最小。反向传播(BP)算法是人工神经网络训练中的常见方法,在训练的过程中, BP算法要计算对网络中所有连接的权重计算损失函数的梯度,根据这一梯度值来更新连接的权值,进而最小化损失函数^[5]。BP算法最早在 20 世纪 70 年代被提出,这一算法在浅层的神经网络训练中取得了重要的成功,然而在面对深度神经网络时,这一算法会遇到“梯度消失问题”,即前面的隐藏层中的神经元的学习速度

收稿日期: 2017-05-28
网络出版日期: 2017-07-06

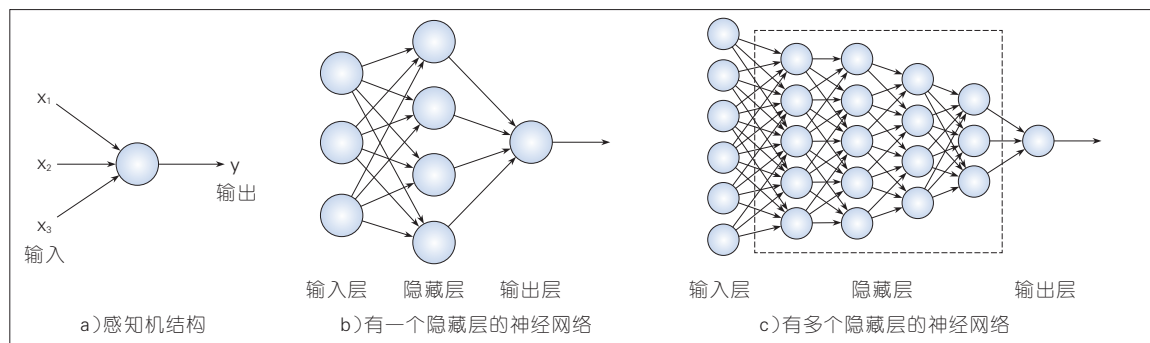


图1
神经网络结构示意图

要慢于后面的隐藏层,这一问题如果无法解决,那么神经网络将在事实上无法有效利用多个隐藏层。这一困难直到2006年才被加拿大多伦多大学教授Geoffrey Hinton解决,这成为了深度学习领域的标志性事件,它使得神经网络和深度学习重新被学术界所重视^[4]。在短短10余年的时间里,深度学习成为了学术界和工业界最为热门的研究主题,在许多不同的领域得到了广泛的应用。深度神经网络也发展出了诸多不同种类的变形。要想真正理解这些不同形式的深度神经网络的工作原理,我们首先必须对这些网络的结构基础进行深入的研究。在文章中,我们将以自动编码器(AE)和限制玻尔兹曼机(RBM)为例,介绍其工作原理和训练方法,在此基础上,我们将讨论这些基本结构在深度学习中的应用。

1 自动编码器

在许多复杂的深度学习问题中,我们都能见到AE的身影。一个AE包括两个基本的组成单元:编码器 f 和解码器 g ,两者本身可以是多层的神经网络,它能将输入端的信号在输出端复现出来。AE为了实现这种复现,就必须提取那些输入数据中最为核心的特征,从而实现有效的复现,对原始输入数据的复现称为一次“重构”。一个AE的结构如图2所示,编码器将输入数据 x 编码到隐藏层 h ,这一编码过程可以用映射表示为 $h = f(x)$,随后解码器将隐藏层的表示解码为输出端的重构结果 r ,这一解

码过程用映射表示即为 $r = g(h)$ 。从概率的角度来看,我们也可以将编码器映射 f 推广为一个编码分布 $p_e(h|x)$,解码器分布可以推广为 $p_d(r|h)$ 。一个好的重构需要尽可能与输入数据相近。因此,学习过程可以简单描述为最小化损失函数 $L(x, g(f(x)))$ 。

长期以来,AE被认为是无监督学习的一种可能的方案,其“编码—解码”过程不是对原始数据的简单重复,在一个编码—解码过程中,我们真正关心的是隐藏层 h 的特性,例如:当隐藏层 h 的数据有着比原始输入数据更低的维度,则说明编码器 f 将复杂的输入数据在隐藏层 h 用较少的特征重现了出来,这一过程不需要外加的其他标签,因而被称为无监督学习^[5]。这与传统的主成分分析(PCA)方法有类似之处,换言之,在线性的情况下,如果我们按照均方误差来定义惩罚函数 L ,此时的AE就是经典的PCA。AE提供了一种更为通用的框架,它可以有更好的非线性特性。AE可以用于进行数据降维、数据压缩、对文字或图像提取主

题并用于信息检索等。

根据所解决的问题不同,AE可以有多种不同形式的变形,例如:去噪自编码器(DAE)、变分自编码器(VAE)、收缩自编码器(CAE)和稀疏自编码器等^[5]。下面我们主要以DAE为例介绍AE的训练过程^[6]。当输入的信息中包含噪声时,输入数据 x' 中包含了真实信息 x 与噪声的叠加,此时,如果考虑最小化损失函数 $L(x, g(f(x')))$,那么我们就得到了一个DAE,其输入为包含噪声的信号 x' ,输出为消除了噪声的信号 x 。考虑一个信息的损坏过程 $C(x'|x)$,该条件分布代表了给定真实样本 x 产生包含噪声的输入数据 x' 的概率。通过最小化损失函数 $L = -\log p_e(x|h=f(x'))$,这一最小化过程可以通过经典的BP算法来进行梯度下降,一个训练好的DAE可以从输入数据 x' 中很好地重构原始数据 x 。DAE学习了输入信号的更鲁棒的表示方式,其泛化能力比一般的AE更强^[1,6]。

多个AE可以逐层堆叠,组合出堆叠式自动编码器(SAE)^[7]。SAE的训练通常采用逐层训练的方法来进行:如图3a)中所示,对于单层的AE,可以通过最小化输入端和经编码—解码得到的输出信号之间的重构误差来对其进行训练;而对于如图3b)所示的多个AE的堆叠,在训练的过程中,可以采用类似于3a)中的训练方法逐层对各层编码器进行训练,如果单层的AE已经被训练好,那么可以认为其编码已经能够较好地重构输入数据。在训练第 n 层编码

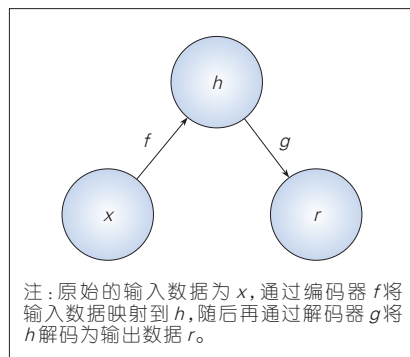
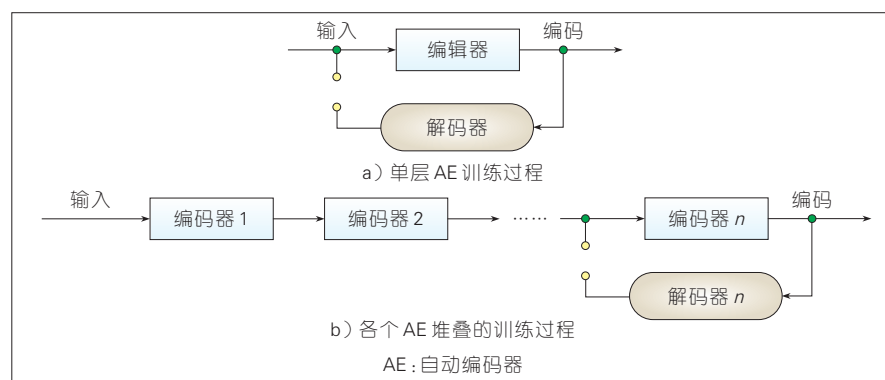


图2 AE的结构



▲图3 AE的训练过程

器时,我们考虑对第 $n-1$ 层的编码结果进行再次编码,最小化输入端(即 $n-1$ 层的编码结果)与经编码—解码得到的输出信号之间的重构误差。在逐层训练的过程中,解码器本身对于训练编码器是重要的,而当整个编码器被逐层训练好之后,解码器本身也就不再需要了。

在具体的应用中,除了直接用AE无监督地提取特征以外,更常见的应用是在一个SAE后再接一个分类器,即用提取出来的特征来对系统的状态进行分类——这就变成了一个监督学习的问题。要解决这样的问题,通常分为2个阶段:在第1阶段进行无监督的预训练,通过逐层训练的方式得到一个容易进行下一步训练的神经网络;随后再进行有监督学习进行微调,最终得到一个能较好地处理分类问题的神经网络。这种逐层训练的方法是深度学习的基础^[4],如果直接训练一个深层的自动编码器,那么常常会由于遇到梯度扩散等问题而导致训练效果不佳,而逐层训练的方法可以有效地避免这些问题。

2 从限制玻尔兹曼机到深度信念网络

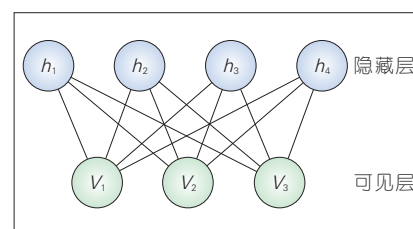
RBM是实现深度学习的另一种神经网络结构。RBM的基本功能与AE类似,RBM同样也可以对数据进行编码,多层RBM也可以用于对神经网络进行初始化的预训练。RBM的基本结构如图4所示,一个基本的

RBM构成了一个二分图,其连接仅在隐藏层和可见层二者之间存在,在隐藏层或可见层的内部并不存在连接,这种特殊连接结构使得RBM的可见层(或隐含层)内各单元间保持了相互独立,这可以简化后续的条件概率计算及采样^[1,8]。

RBM是一种基于能量的神经网络模型,它具有深刻的统计物理背景。在统计物理学中,由于系统倾向于处在能量较低的状态,因此如果某一种状态的能量越低,系统就会有越大的概率处在这一状态。在平衡态的情况下,在服从玻尔兹曼分布时,系统出现某一状态 s 的概率 P 与这一状态的能量 E 的负指数成正比,即 $P \sim e^{-E}$ 。统计物理中能量与概率分布之间的关系启发我们将概率分布转变成基于能量的模型。此时,系统中复杂而抽象的各种状态出现的概率也就被用一个简单的能量函数给表示出来了。对于一个有 n 个隐含层节点和 m 个输入层节点的RBM,其能量函数 $E(v, h)$ 是这样定义的:

$$E(v, h) = -\sum_{i=1}^n \sum_{j=1}^m w_{ij} h_i v_j - \sum_{j=1}^m b_j v_j - \sum_{i=1}^n c_i h_i \quad (1)$$

式(1)中的各 w_{ij} 构成了RBM的连接矩阵,而 b_j 和 c_i 则分别表示了可见层和隐藏层的偏置,它们构成了模型的参数,而 v_j 和 h_i 则是对系统状态的刻画。这个函数与物理学中经典的Ising模型的能量函数是类似的^[9]。有了能量函数,系统处在各状态的概率



▲图4 RBM的基本结构示意图

分布及各种边缘分布则可得到。

RBM训练的目标即为让RBM网络表示的概率分布与输入样本的分布尽可能地接近,这一训练同样是无监督式的。在实践中,常常用对比散度(CD)的方法来对网络进行训练,CD算法较好地解决了RBM学习效率的问题^[10]。在用CD算法开始进行训练时,所有可见神经元的初始状态被设置成某一个训练样本,将这些初始参数代入到激活函数中,可以算出所有隐藏层神经元的状态,进而用激活函数产生可见层的一个重构^[10]。通过比照重构结果和初始状态,RBM的各参数可以得以更新,从这一点来看,用CD算法对RBM进行训练与AE的训练是非常相似的。近年来,CD算法有许多改进,例如:持续性对比散度(PCD)和快速持续性对比散度(FPCD)等;而在训练RBM时,也可以利用非CD式的算法,例如比率匹配等^[11]。

在RBM的基础上,深度信念网络(DBN)由Hinton等人提出,DBN的工作开启了深度学习的序章^[1,2,8]。在传统的BP神经网络中,一旦网络的深度增加,训练的过程将会变慢,同时不适当的参数选择将导致网络收敛于局部最优。此外,在训练时还常常需要为训练提供一个有标签的数据集,而DBN则很好地解决了这一问题。与我们提到的SAE类似,也可以通过叠加RBM来建立DBN。第1个RBM是整个网络的输入层,第 l 层RBM的隐藏层作为第 $l+1$ 个RBM的可见层。在训练DBN时,首先用CD算法来训练第1个RBM,然后将其隐藏层作为第2个RBM的可见层,用这

些数据来对第2个RBM用CD算法进行训练,以此类推。最终,当整个网络的所有层都被训练完之后,整个DBN也就被预训练好了。这样一个训练过程是无监督的学习过程,各层RBM都在尽可能反映其上一层数据的特征,这一过程与SAE的训练过程类似,是对网络各参数的初始化。在DBN的最后一层也可以再接一个分类器,进行有监督学习,用BP算法微调整个DBN,这种方法使得DBN克服了直接对深度神经网络进行训练时易出现的局部最优等问题,使得深层神经网络真正有了应用价值^[8,11]。

3 自动编码器与限制玻尔兹曼机的区别

AE与RBM两种算法之间也有着重要的区别,这种区别的核心在于:AE以及SAE希望通过非线性变换找到输入数据的特征表示,它是某种确定论性的模型;而RBM以及DBN的训练则是围绕概率分布进行的,它通过输入数据的概率分布(能量函数)来提取高层表示,它是某种概率论性的模型。从结构的角度看,AE的编码器和解码器都可以是多层的神经网络,而通常我们所说的RBM只是一种两层的神经网络。在训练AE的过程中,当输出的结果可以完全重构输入数据时,损失函数 L 被最小化,而损失函数常常被定义为输出与输出之间的某种偏差(例如均方差等),它的偏导数便于直接计算,因此可以用传统的BP算法进行优化。RBM最显著的特点在于其用物理学中的能量概念重新描述了概率分布,它的学习算法基于最大似然,网络能量函数的偏导不能直接计算,而需要用统计的方法进行估计,因此需要用CD算法等来对RBM进行训练。

4 结束语

深度学习是机器学习领域中的重要算法,它通过构建多个隐含层的神经网络和海量的训练数据,来提取

更有用的特征,并最终提升分类或预测的准确性。AE和RBM是深度学习的结构基础,它们本身都可以作为一种无监督学习的框架,通过最小化重构误差,提取系统的重要特征;更重要的是,通过多层的堆叠和逐层的预训练,SAE和DBN都可以为后续的监督学习提供一个良好的初值,从而让神经网络可以更好更快地达到收敛^[11]。正是这种重要的性质使得深度学习在过去10余年的发展中取得了重要的成功。深度学习能解决的问题变得越来越复杂,同时其精度不断提高。

深度学习是一个人工智能领域迅速发展的方向,随着计算能力的提高以及规模更大的数据集合的出现,深度学习的规模也在不断增长,深层神经网络的优势也在不断体现——许多曾经被认为较为抽象、难以分类的复杂特征在深度学习的框架下也变成了可以解决的问题,这使得深度学习算法在许多不同的领域都发挥了重要的应用,并且还有着广阔的应用前景^[1,2,12]。更重要的是,深度学习还为解决大量的无监督学习问题提供了可能性。从无标签的数据中进行无监督学习一直以来都是研究人员所面临的一个主要挑战,在这方面,深度学习仍远远无法与人类智能相比,不过近年来,在无监督学习领域也已经出现了许多重要的突破。相信在不久的将来,我们将看到越来越多深度学习和人工智能领域的重大突破,也将看到相关算法在许多新领域的应用和机遇。

参考文献

- [1] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. USA:MIT Press, 2016.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep Learning[J]. Nature, 2015, 521(7553): 436–444. DOI:10.1038/nature.14539
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks[C]//Advances in Neural Information Processing Systems. USA: Neural Information Processing Systems Foundation, 2012: 1097–1105
- [4] HINTON G E, SALAKHUTDINOV R R.

Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313 (5786): 504–507

- [5] BALDI P. Autoencoders, Unsupervised Learning, and Deep architectures[J]. ICML Unsupervised and Transfer Learning, 2012, 27(37–50): 1
- [6] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and Composing Robust Features with Denoising Autoencoders [C]// Proceedings of the 25th International Conference on Machine Learning. USA: ACM, 2008: 1096–1103
- [7] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. Journal of Machine Learning Research, 2010 (11): 3371–3408
- [8] HINTON G E, OSINDERO S, TEHY W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7): 1527–1554
- [9] MEHTA P, SCHWAB D J. An Exact Mapping Between the Variational Renormalization Group and Deep Learning[J]. DOI:1410.3831, 2014
- [10] HINTON G E. Training Products of Experts by Minimizing Contrastive Divergence[J]. Neural Computation, 2002, 14(8): 1771–1800
- [11] ERHAN D, BENGIO Y, COURVILLE A, et al. Why Does Unsupervised Pre-Training Help Deep Learning?[J]. Journal of Machine Learning Research, 2010(11): 625–660
- [12] SCHMIDHUBER J. Deep Learning in Neural Networks: An Overview[J]. Neural networks, 2015(61): 85–117

作者简介



康文斌, 湖北医药学院教师, 湖北医药学院Bio-X研究中心主要负责人;研究方向为软凝聚态物理与生物物理、蛋白质物理学问题以及机器学习算法在药物设计等领域中的应用;已发表2篇SCI论文, 2篇EI论文以及3篇其他期刊论文。



彭菁, 平安科技(深圳)有限公司数据分析师;研究方向为无监督学习在用户行为数据分析中的应用, 尤其关注用户在各金融领域(保险、信贷、理财等)行为轨迹研究以及相关问题的商业价值探索。



唐乾元, 香港浸会大学物理系博士后;研究方向为统计物理在生命科学和人工智能领域中的应用, 包括深度学习算法的统计物理基础以及深度学习在生物大数据分析中的应用;已发表SCI论文2篇。

深度学习进展及其在图像处理领域的应用

Advances in Deep Learning and Its Applications in Image Processing

刘涵/LIU Han¹
贺霖/HE Lin²
李军/LI Jun¹

(1. 中山大学, 广东 广州 510275;
2. 华南理工大学, 广东 广州 510640)
(1. Sun Yat-Sen University, Guangzhou
510275, China;
2. South China University of Technology,
Guangzhou 510640, China)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0036-005

摘要: 深度学习一般通过3种方式进行: 有监督学习、无监督学习和混合深度学习。以“无监督或生成式特征学习”以及“有监督特征学习和分类”为例, 讨论了深度学习及其在图像处理等领域的进展及未来可能的研究方向。认为深度学习打破了传统机器学习和信号处理技术普遍基于浅层结构的局限。得益于相关非凸优化等问题的逐步解决, 深度学习已经在图像处理等领域取得了一些突破性的进展。

关键词: 深度学习; 图像处理; 分层结构

Abstract: Deep learning methods are usually divided into three different categories, including supervised learning, unsupervised learning and mixed-structured learning. In this paper, "unsupervised or generative feature learning" and "supervised feature learning and classification" are taken for examples to illustrate the research advances in deep learning, and several promising research lines in image processing. Deep learning technique has overcome some limitations of shallow structures commonly used in traditional machine learning and signal processing. Due to the progress in some aspects, such as relevant nonconvex optimization, deep learning has achieved remarkable development.

Keywords: deep learning; image processing; hierarchical structure

在过去10年左右的时间里, 深度学习对信息技术的许多方面都产生了重要影响。诸多关于深度学习的描述普遍存在两个重要的共同点: 包含多层或多阶非线性信息处理的模型; 使用了连续的更高、更抽象层中的监督或无监督学习特征表示的方法。深度学习是以神经网络为基础, 包含人工智能、图模型、最优化等技术在内的交叉领域。它之所以如此受关注, 主要源于3个方面: 芯片硬件处理性能的巨大提升, 为深度网络的复杂计算提供了基础; 用于训练的数据呈爆炸性增长, 为复杂网络的学习提供了可能; 机器学习和信息处理等方面研究取得了很大进展。

1 深度学习的发展

以前, 绝大多数机器学习和信号处理技术都是基于浅层结构, 如高斯

混合模型(GMM)、线性或非线性动力系统、条件随机场(CRF)、最大熵模型(MaxEnt)、支持向量机(SVM)、逻辑回归(LR)、核回归以及多层感知器(MLP)等。这些结构一般包含最多一到两层的非线性特征变换。已有研究表明: 浅层结构在解决简单的或者约束较多的问题上效果明显, 但是由于其建模和表示能力有限, 在对实际应用中一些较为复杂自然信号(比如人类语音、自然声音和语言、自然图像和视觉景色)进行处理时会遇到一些困难。人类的听觉和视觉信息等的处理机制一般可以用深度结构描述, 通过该结构可以从感官输入信息中提取复杂结构并构建内部

表示。如果能实现有效和高效的深度学习算法, 那么对于各种自然信号的处理技术而言, 其性能会得到很大提升。

深度学习的概念一般被认为来源于对人工神经网络的研究。前馈神经网络或具有多隐层的多层感知器是深度神经网络(DNN)的典型模型。反向传播(BP)算法是解决其学习问题的广泛运用的典型算法。遗憾的是, 仅仅使用BP算法在实际学习隐层数目较多的网络时往往效果不是很好^[1]。在优化目标为非凸函数的深度神经网络学习中, 通常存在局部最优解等问题。BP算法基于局部梯度信息, 往往从一些随机的初始点

收稿日期: 2017-05-23

网络出版日期: 2017-07-12

基金项目: 国家自然科学基金(61571195); 广东省自然科学基金(2016A030313516)

开始寻优,当使用批量梯度下降或随机梯度下降的BP算法时,目标函数经常会陷入局部最优。随着网络层数的加深,局部最优的情况也就会变得越来越严重。虽然相关研究者对小规模的神经网络的探究从未间断过,但是在很多机器学习和信号处理方法中,研究者们将研究重点从对神经网络本身的研究转移到对具有凸损失函数的浅层模型的研究,这些模型以降低建模准确度为代价,达到快速高效地收敛到全局最优化的目的。所以,深层网络本身还存在着易于陷入局部最优等缺陷,有待于进行更深入的研究。

Hinton 等在 2006 年左右提出了一种高效的基于深度置信网络 (DBN) 的无监督学习算法^[2-3],他们利用经验性的方法处理了与深度模型相关的最优化难题。DBN 是一种深度生成式模型,由一组受限玻尔兹曼机 (RBMs) 堆叠而成,它的核心部分是贪婪式的逐层学习,这种算法可以最优化 DBN 中的权重,且其时间复杂度与网络的大小和深度呈线性关系。最近,相关研究者对于 DNN 与 DBN 进行了更加细致的研究,如可使用 DBN 来初始化 DNN 的权值等。在 DNN 中,多隐层的使用不仅显著提高了网络的表示能力,而且可得到一些较优解。然而,在训练过程中使用深而宽的神经网络需要依赖于强大的计算性能。随机梯度下降 (SGD) 算法就是一种在训练集较大且冗余的情况下较为有效的学习算法^[4]。已有的研究表明:SGD 可以有效地实现并行方式的运算。该并行运算主要通过两种方式实现:一种方式是通过异步模式使用多台计算机^[5];另一种方式是使用多图形处理器 (GPU) 的流水线型的 BP 算法^[6]。另外,从单个或小批量样本中估计得到的随机性梯度使得 SGD 通常能跳出局部最优解。其他的一些学习算法,如 Hessian free^[7]、Krylov subspace^[12] 方法等,都表现出了类似的学习能力。对

于 DNN 学习中涉及的非凸优化问题,更好的参数初始化和学习技术都会学习出更好的模型。

DBN 预训练并不是唯一可对 DNN 进行有效初始化的方法。基于降噪自动编码器的方法对 DNN 进行逐层地预训练,将每两层视为一个降噪自编码器,该编码器再通过将输入节点的随机子集设置为零进行正则化^[13]。另一种方法则是使用压缩自编码器^[14],该编码器通过使输入变量具有更好的鲁棒性来达到同样的目的。此外,Ranzato 等开发了稀疏编码对称机 (SESM)^[9],其在构建 DBN 模块中具有和 RBM 非常类似的架构,它也可以用来有效地初始化 DNN 的训练过程。除了使用贪婪方法逐层地进行无监督预训练,有监督的预训练 (有时称为判别式预训练) 也被证明是比较有效的^[10-11]。有监督的预训练的基本思路是从一个经过 BP 算法训练的单个隐层 MLP 开始,每一次需要添加一个新的隐层时,用一个随机初始化的新的隐层和输出层替换输出层,并用 BP 算法训练全新的 MLP (或 DNN)。在 RBM 发展的同时,出现了另外两种较有代表性的非概率的、非生成式的深度模型:一种是基于自编码器 (AE) 的改进模型,其使用与 DBN 训练相似的贪婪分层方法进行训练;另一种是基于能量的模型,其利用稀疏表示来进行非监督学习。与 DBN 相似,其也可对深度神经网络进行高效的预训练。

2 3 种深度学习网络

深度学习是一类应用广泛的机器学习技术和架构,其特点是采用多层的非线性结构进行信息处理,这种方法在本质上是分层实现的。根据不同应用领域的任务目标及对应的不同深度学习网络结构,我们可以大致把已有深度学习结构分为 3 类:

(1) 无监督的或生成式学习的深度网络。该结构针对模式分析和合成任务,用于在没有目标类标签信息

的情况下捕捉可见数据的高阶相关性。已有的无监督特征学习或表达学习指的就是这一类深度网络。

(2) 有监督学习的深度网络。该种网络直接提供用于模式分类目的的判别能力,它的特点是描述了在给定可见数据的条件下不同类别的后验概率分布。对于这种有监督的学习,目标数据的类别标签总是以直接或间接形式给出,所以它们也被称作判别式深度网络。

(3) 混合式深度网络。其目标是实现判别式模型的效果,往往以生成式或无监督深度网络的结果作为重要辅助,通过更好地优化和正则化以上类别 (2) 中的深度网络来实现,也可以通过使用判别式准则对以上类别 (1) 中所述的深度生成式或无监督深度网络进行参数估计来实现。

从传统机器学习的角度,深度学习模型可分为深度判别式模型和生成式/无监督模型。然而这种模型分类方法忽略了深度学习研究中的一个重要观点,即生成式和无监督学习模型可通过更好地正则化与优化来提高深度判别网络的训练效果。因此,深度学习网络有时会以混合式的结构形式出现。

3 深度学习在图像目标识别和计算机视觉中的应用

多年来,计算机视觉和图像目标识别等任务长期依赖人工设计的特征,如尺度不变特征变换 (SIFT) 和方向梯度直方图 (HOG) 等。此类特征仅仅是对图像中低级别的边缘信息进行描述与表征,若要描述图像中高级信息例如边缘交叉和局部外观等,其往往显得力不从心。深度学习可以通过无监督和有监督的学习方法直接从数据中获得层级化的视觉特征,从而提供一套更为有效的解决方案。深度学习方法经常可从无监督和有监督两个角度进行讨论:无监督特征学习,该类方法通常将深度学习用于特征提取,然后这些特征会被直

接送入后续分类算法;有监督的特征学习,当存在大量有标签样本时,此类方法通过端到端的学习策略实现特征提取与分类器的联合优化。

3.1 无监督或生成特征学习

当有标签样本相对缺乏时,无监督学习算法可用于学习视觉特征层级结构。如 Hinton 和 Salakhutdinov 等最早提出将无监督深度自编码方法应用于 DBN 模型的预训练^[4]。他们利用该方法在仅有 60 000 个训练样本的美国国家标准技术研究所修正(MNIST)数据库上成功实现了图像的识别和降维(编码)任务。此外, Nair 和 Hinton 提出了一种改进的 DBN,该 DBN 的顶层使用了一个三阶的 RBM^[12]。当这种 DBN 被应用于 NORB 数据库(一个三维目标识别任务数据库)上时,其错误率几乎下降到了目前所公布的最低水平,这再次表明了 DBN 在很大程度上是优于类 SVM 这样的浅层模型。随后,还出现一些对 DBN 的改进文献。其他的一些无监督深度特征学习方法还包括稀疏自编码器及基于深度稀疏编码的模型等^[13]。

3.2 有监督特征学习和分类

CNN 是一种受到广泛关注的有监督深度学习结构。有监督 CNN 结构获得广泛关注始于 2012 年 10 月 ImageNet 竞赛,这主要是由于大量的有标签样本及高性能 GPU 计算平台的出现使得大规模 CNN 的高效训练成为可能。图 1 给出了文献^[14]中所描

述的 CNN 的基本结构。为了实现典型图像像素的空间位置相对不变性这一特点,CNN 使用了一个带有局部连接和共享权值的卷积层,该层的输出通过一个非线性激活函数来获得激活响应,接着通过一个非线性池化层来减小数据量,最后再将池化层的输出连接到若干个全连接层。这种结构也常被称作深度卷积神经网络。

CNN 在 2012 年的 ImageNet 竞赛中取得了瞩目的成绩。在该次比赛中,使用深度 CNN 进行建模的方法获得了前所未有的低错误率。该深度 CNN 模型包含 6 千万个权值,65 万个神经元节点以及结合 5 个卷积层的最大池化层。此外,两个全连接层也被用于这个 CNN 模型的最顶层。另外,还有两个额外的因素也起到了很重要的作用:首先,是一个称为“dropout”^[15]的强大的正则化技术;第二个重要因素是通过激活函数 $f(x) = \max(x, 0)$ 所实现的整流线性单元(ReLU)的应用,使得整个训练过程的效率被极大地提高,尤其是通过 GPU 并行运算实现后取得的效果更加明显。其后,基于更大规模的模型以及更多的训练数据,CNN 得到了进一步的改进。不少深度 CNN 模型和方法的强大学习能力在各年的 ImageNet 竞赛上得到了验证。

深度 CNN 已被证明在图像目标识别任务中具有卓越的分类性能,关于其机理也逐渐有了一些解释,如 Zeiler 等利用基于反卷积网络的可视化技术对 CNN 的机理进行了一些讨论^[16]。图 2 示意了反卷积机理,该反

卷积网络通过 CNN 中相应前馈计算的相反连续操作,其中包括反池化、校正和滤波,使得特征图谱之上的活动得以重建。在实现反池化的过程中,最大池化操作的非可逆性通过近似逆向逼近的方法得以解决。

除了深度 CNN 结构外,DNN 结构也在大量的计算机视觉任务上获得了成功^[17-19]。目前,基于深度 CNN 结构的有监督学习模式及其相关的分类技术已在相关研究领域产生很大影响,这尤其体现在 2012—2013 年的 ImageNet 比赛中。这些方法不仅可以用于图像目标识别任务,同样还可以应用于其他一些计算机视觉的任务中。当然,关于 CNN 深度学习方法的机理及其局限性等,仍有很多问题需要探讨。

4 结论及展望

深度学习作为当前机器学习领域最热门的技术之一,已经在不少领域获得了应用,并且展现出巨大的前景。根据网络结构的不同,它大致可以分为无监督、有监督和混合神经网络 3 种类别。作为一种从本质上来说是分层非线性结构的深度模型,它所构建和学习的深层特征表示无疑极大地提升了传统浅层模型的泛化能力;但是随之而来的明显非凸的优化目标形式却长期困扰着该领域的研究者,如其导致的局部最优化等问题,阻碍着基于梯度的 BP 算法的有效实施。得益于无监督预训练方法对优化初始点的改善,该问题有了一些经验性且行之有效的解决方案。

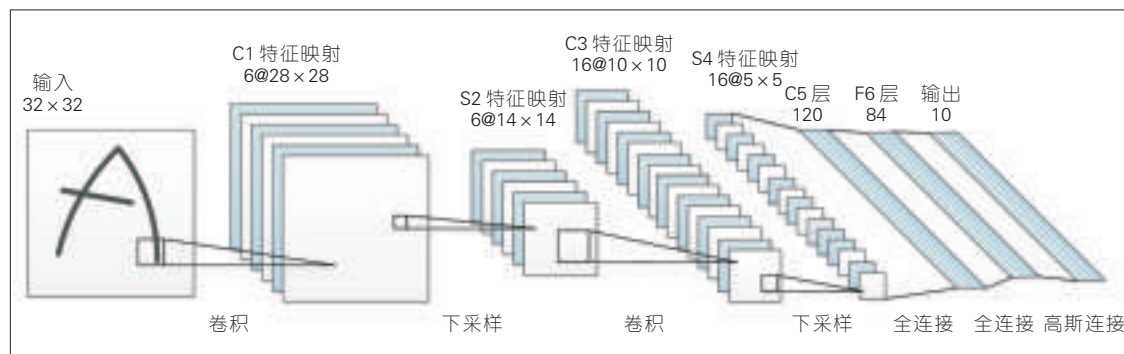
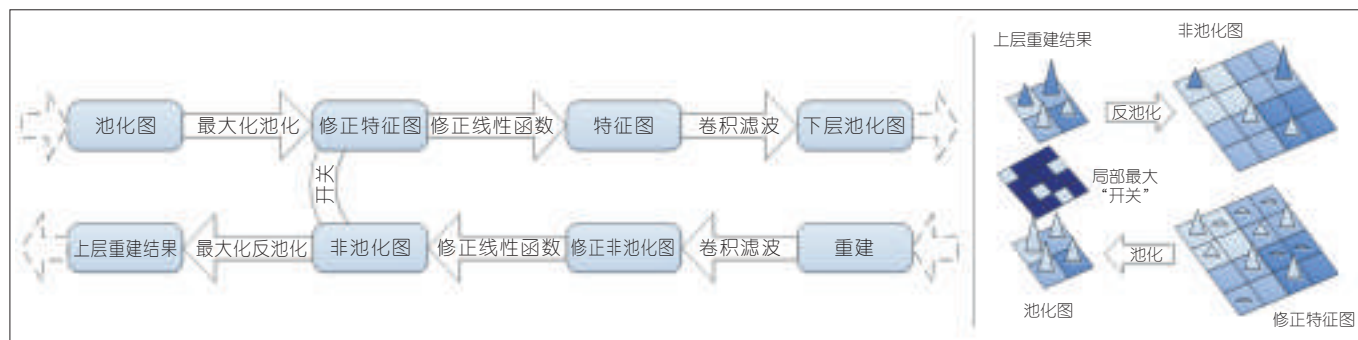


图1
一种典型的CNN结构



▲图2 反卷积网络机理

虽然该类方法缺乏稳固的理论基础,却成功促成了深度学习方法在学术界和工业界的大规模成功应用。通过最近的研究发现,现有深度架构在优化技术等方面存在着巨大的提升空间^[7,10,20-23]。

另一方面,如果训练数据集足够大,理论上讲,模型的泛化能力将会得到较大的提升,那么通过深度置信网络等预训练方法所带来的良好优化初始点的重要性必然会显著降低。然而,要实施针对大规模数据集的应用,强大的计算能力是必不可少的。当前来看,有效且可拓展的并行算法是训练庞大数据集的关键所在。然而,常用的基于 mini-batch 的梯度下降技术很难并行实现。最近出现的异步梯度下降等技术为这一领域相关问题的解决带来了一些新思路,并在 CPU 集群^[7,13]和 GPU 集群^[24]中得到了初步的实现。未来,对于并行学习和新型的大规模优化算法仍需要进行有针对性的理论研究。

目前,阻碍深度模型发展的另一主要问题在于超参数的合理选择。众所周知,基于神经网络的深度学习技术有着数量众多且自由度极大的超参数,如网络架构的层数以及每层的单元数、正则化强度、学习速率以及学习速率衰减率等。基于传统的网格搜索等技术的解决方案无论从效率还是成本的角度上来讲对于超参数的设定都是不可行的。此外,不同的超参数之间通常存在着相互依赖性,且微调代价巨大。这些问题决

定了我们需要开展进一步的研究来探索更有效的解决方案。令人欣慰的是,近来以随机采样^[25]和贝叶斯优化过程^[26]为代表的方法给我们带来了一些新思路。虽然一些经验性的超参数设置即可在一定程度反映出深层结构所具有的强大泛化能力,但是超参数的最优化配置在对一些特定应用方面进一步提高深度模型性能上仍具有很大的研究价值。

领域知识的应用是深度学习方法成功的另一大关键因素。根据不同任务的特点设计不同的不变性特征提取方法以及正则化方法是当前所流行的一种应用方式。研究者们也在基于域适应的迁移学习对于深度学习的辅助作用方面进行了一些研究,但是能够普遍适用于各类分类任务的深度学习技术仍然是不存在的,例如:当前较为通用的生成式预训练伴随判别式微调的学习策略在一些特定的任务(如语音识别)中表现并不理想。因此,更加有效的域适应技术和新型的通用学习架构对于图像处理等领域中的一些较为复杂的问题而言是极其重要的。

深度学习理论还有其他一些方面的基础性理论问题值得关注,例如:如何通过关注数据变化中潜在因素的分分布式表示问题来设计更合理的深度学习架构,进而提取更高效的特征表示;在深度结构的输入输出表示中同时引入结构信息^[27-29],使得大多数传统的深度学习技术只能用于“扁平结构”表示的缺陷得到改善。

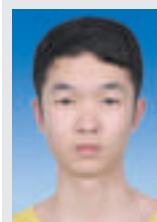
最后,为了实现理想中的“强人工智能”,从而实现具备类似人类大脑水平的智慧,传统的以信号处理和机器学习为主要技术基础的人工智能研究应更多地寻求同神经计算等领域的合作,通过借助于前沿生物领域对人类大脑分层结构的最新研究成果来改善当前的系统计算模型。

参考文献

- [1] BENGIO Y. Learning Deep Architectures for AI [J]. Foundations and Trends® in Machine Learning, 2009, 2(1):1-127. DOI: 10.1561/2200000006
- [2] HINTON G E, OSINDRO S, TEH Y. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1527-1554. DOI: 10.1162/neco.2006.18.7.1527
- [3] HINTON G E, SALAKHUTDINOV R R. Reducing the Dimensionality of Data with Neural Networks[J]. science, 2006, 313(5786): 504-507. DOI: 10.1126/science.1127647
- [4] BOTTOU L, CUN Y L. Large Scale online Learning[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2004
- [5] DEAN J, CORRADO G, MONGA R, et al. Large Scale Distributed Deep Networks[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2004
- [6] CHEN X, EVERSOLE A, LI G, et al. Pipelined Back-Propagation for Context-Dependent Deep Neural Networks[C]//Interspeech 2012. USA: IEEE, 2012. DOI: 10.1.1.649.218
- [7] MARTENS J. Deep Learning via Hessian-Free Optimization[C]//Proceedings of the 27th International Conference on Machine Learning (ICML-10). USA: IEEE, 2010. DOI: 10.1.1.170.2846
- [8] VINCENT. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. Journal of Machine Learning Research, 11(Dec): 3371-3408. DOI: 10.1561/2200000006

- [9] BOUREAU Y, CUN Y L. Sparse Feature Learning for Deep Belief Networks[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2008
- [10] BENGIO Y. Greedy Layer-Wise Training of Deep Networks[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2007:153
- [11] YU K, LIN Y. Learning Image Representations from the Pixel Level via Hierarchical Sparse Coding[C]//2011 IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2011. DOI: 10.1109/CVPR.2011.5995732
- [12] MNH V, KAVUKCUOGLU, SILVER D, et al, Playing Atari with Deep Reinforcement Learning[EB/OL].(2017-05-22). <https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf>
- [13] LE Q V. Building High-Level Features Using Large Scale Unsupervised Learning[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. USA: IEEE, 2013. DOI: 10.1109/ICASSP.2013.6639343
- [14] LE C Y. Gradient-Based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998. 86(11): 2278-2324. DOI: 10.1109/5.726791
- [15] HINTON G E, SRIVASTAVA N, KRIZHEVSKY, et al, Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors[EB/OL].(2017-05-22). <https://arxiv.org/pdf/1207.0580.pdf>
- [16] ZEILER M D, FERGUS R. Visualizing and Understanding Convolutional Networks[C]//European Conference on Computer Vision. EU: Springer, 2014
- [17] CIRESAN D C. Deep, Big, Simple Neural Nets for Handwritten Digit Recognition[J]. Neural Computation, 2010, 22(12): 3207-3220. DOI: 10.1162/NECO_a_00052
- [18] CIRESAN D. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2012
- [19] CIREGAN D and MEIER U. Multi-Column Deep Neural Networks for Image Classification[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2012. DOI:10.1.1.367.484
- [20] MARTENS J, SUTSKEVER I. Learning Recurrent Neural Networks with Hessian-Free Optimization[C]//The 28th International Conference on Machine Learning. USA: IEEE, 2011
- [21] LE V Q, NGUAN J, COATES A, et al. On Optimization Methods for Deep Learning[C]//The 28th International Conference on Machine Learning. USA: IEEE, 2011
- [22] SAINATH T N. Optimization Techniques to Improve Training Speed of Deep Neural Networks for Large Speech Tasks[J]. IEEE Transactions on Audio, Speech, and Language Processing, 21(11): 2267-2276. DOI: 10.1109/TASL.2013.2284378
- [23] WRIGHT S J. Optimization Algorithms and Applications for Speech and Language Processing[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(11): 2231-2243. DOI: 10.1109/TASL.2013.2283777
- [24] COATES A, HUVAL B, WANG T, et al. Deep Learning with COTS HPC Systems[C]//International Conference on Machine Learning. USA: IEEE, 2013
- [25] BERGSTRÄ J, BENGIO Y. Random Search for Hyper-Parameter Optimization [J]. Journal of Machine Learning Research, 2012, 13: 281-305. DOI: 10.2307/1268522
- [26] SNOEK J, LAROCHELLE H, ADAMS P R. Practical Bayesian Optimization of Machine Learning Algorithms[C]//Advances in Neural Information Processing Systems. USA: IEEE, 2012
- [27] SOCHER R. New Directions in Deep Learning: Structured Models, Tasks, and Datasets[C]//Neural Information Processing Systems. USA: NIPS, 2012
- [28] DENG L. Design and Learning of Output Representations for Speech Recognition[C]//Neural Information Processing Systems. USA: NIPS, 2013
- [29] SRIVASTAVA N, SALAKHUTDINOV R R. Discriminative Transfer Learning with Tree-Based Priors[C]//Advances in Neural Information Processing Systems. USA: NIPS, 2013

作者简介



刘涵, 中山大学地理科学与规划学院学生; 主要研究领域为遥感图像分类及混合像元分解等。



贺霖, 华南理工大学副教授; 主要研究领域为统计模式识别、统计信号处理和多维图像处理等; 主持国家级及省部级科研项目多项。



李军, 中山大学教授; 主要研究领域为多维图像处理 and 机器学习等; 主持国家级及省部级科研项目多项; 国家“青年千人计划”入选者。

本期专家论坛
栏目策划人

钟义信

北京邮电大学智能科学技术中心教授, 发展中世界工程技术科学院院士, 国际信息研究学会名誉主席及中国分会主席, 曾任国家“863”计划通信主题首届首席专家, 国务院信息化专家委员会常委, 中国人工智能学会理事长, 中国神经网络委员会主席; 长期从事信息科学和人工智能基础理论的研究和教学; 先后创立“全信息理论”“语义信息论”和“机制主义人工智能理论”, 获多项国家级和部级科技奖励; 已发表信息科学及人工智能领域学术著作 16 部, 学术论文 510 余篇。

深度学习的能与不能

What Deep Learning Can or Cannot

于剑/YU Jian

(北京交通大学, 北京 100044)
(Beijing Jiaotong University, Beijing 100044, China)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0041-003

摘要: 深度学习技术的应用日渐广泛, 在语音、图像、文本处理、搜索引擎、广告推荐等领域都取得了巨大的成功。认为深度学习自身具有盲点, 无法解决全部的机器学习问题, 并指出了深度学习的优缺点, 为深度学习的使用者提供了一定的理论指导。最后, 还展望了深度学习的未来发展趋势。

关键词: 深度学习; 傻瓜型学习算法; 专家型学习算法; 白箱算法; 黑箱算法; 相关性; 因果性

Abstract: Deep learning is now widely used and has achieved huge successes in many fields, such as speech, image, natural language processing, search engine and advertising recommendation. In this paper, considering the achilles' heel of deep learning, we propose that deep learning cannot solve all problems in machine learning discipline. Moreover, the advantages and the disadvantages of deep learning are illustrated, and some suggestions for users are proposed. Finally, the future trend of deep learning is pointed out.

Keywords: deep learning; fool-type learning algorithm; expert-type learning algorithm; white box algorithm; black box algorithm; relevance; causality

机器学习作为单独的研究方向, 应该说是 20 世纪 80 年代第 1 届国际机器学习大会 (ICML) 召开之后才有的事情。机器学习存在很多不同的定义, 常用的有 3 个: 第 1 个常用的机器学习定义是“计算机系统能够利用经验提高自身的性能”^[1], 第 2 个常见定义是“学习就是一个基于经验数据的函数估计问题”^[2], 第 3 个常见的机器学习定义是“提取重要模

式、趋势, 并理解数据, 即从数据中学习”^[3]。这 3 个常见定义各有侧重: 第 1 个聚焦学习效果, 第 2 个亮点是给出了可操作的学习定义, 第 3 个突出了学习任务的分类。

虽然机器学习的定义晚至 20 世纪才出现, 但是广义上来说, 机器学习任务, 或者说学习任务, 在人类出现伊始就已有之。在日常生活中, 人们每天都面临如何从自己采集的数据中提取知识进行使用的问题。比如: 大的方面, 需要观察环境的变化来学习如何制定政策使得地球可持

收稿时间: 2017-05-27
网络出版时间: 2017-07-05

续发展;小的方面,需要根据生活的经验买到一个可口的柚子或者西瓜,选择一个靠谱的理发师等。在计算机出现以前,数据采集都是人直接感知或者操作的,采集到的数据量较小,人可以直接从数据中提取知识,并不需要机器学习。如:对于回归问题,高斯在19世纪早期(1809)就发表了最小二乘法;对于数据降维问题,卡尔皮尔逊在1901年就发明了主成分分析(PCA);对于聚类问题,K-means 算法最早也可追踪到1953年^[4]。但是这些算法和问题被归入机器学习,也只有在机器收集数据能力越来越成熟,导致人类直接从数据中提取知识成为不可能之后才变得没有异议。

在过去的30年间,机器学习从处理仅包含上百个样本数据的玩具问题起步,一直发展到今天,已经成为了从科学研究到商业应用的标准数据分析工具,机器学习的研究热点也几经变迁。

1 机器学习发展简史

机器学习最早的目标是从数据中发现可以解释的知识,在追求算法性能的同时,强调算法的解释性。早期的线性感知器、决策树和最近邻等算法可以说是这方面的典型代表作。但是,1969年Minsky指出线性感知器算法不能解决异或问题^[5]。由于现实世界的问题大多是非线性问题,而异或问题又可以说是非线性问题中最简单的问题,由此可以推断线性感知器算法的实际用处不大。这对于以线性感知器算法为代表的神经网络研究而言可以说是致命一击,直接导致了神经网络甚至人工智能的第1个冬天。感知器算法的发明人、神经网络先驱Rosenblatt于1971年因故去世,则更加增添了这个冬天的寒意。

需要指出的是,很多实际应用并不要求算法具有可解释性,比如:机器翻译、天气预报、打卦算命等。在

这种需求下,如果1个算法的泛化性能能够超过其他同类算法,即使该算法缺少解释性,则该算法依然是优秀的学习算法。20世纪80年代神经网络的复苏,其基本思路即为放弃解释性,一心提高算法的泛化性能。神经网络放弃解释性的最重要标志是其激活函数不再使用线性函数,而是典型的非线性函数,如Sigmoid函数和双曲函数等,其优点是表示能力大幅提高,但相应的复杂性也极度增长。众所周知,解释性能好的学习算法,其泛化性能也要满足实际需求,如果其泛化性能不佳,即使解释性好,人们也不会选用。在20世纪80年代,3层神经网络的性能超过了当时的分类算法,如:决策树、最近邻等,虽然其解释性不佳,神经网络依然成为当时最流行的机器学习模型。在神经网络放弃解释性之后,其对于算法设计者的知识储备要求也放到了最低,因此神经网络在20世纪80年代吸引了大批的研究者。

当然,也有很多实际应用要求算法具有可解释性,如因果关系发现、控制等。应该说,同时追求解释性和泛化性能一直是非神经网络机器学习研究者设计学习算法的基本约束。一旦某算法既具有很好的解释性,其性能又超过神经网络,神经网络研究就将面临极大的困境,这样的事情在历史上也曾真实地发生过。1995年Vapnik提出了支持向量机分类算法,该算法解释性好,其分类性能也超过了当时常见的3层神经网络,尤其需要指出的是,其理论的分类错误率可以通过Valiant的概率近似正确(PAC)理论来估计。这导致了神经网络研究的10年沉寂,有人也将其称为人工智能的第2次冬天。在这期间,大批原先的神经网络研究者纷纷转向离开,只有少数人坚持研究神经网络。这个时间段对于机器学习来说,显然不是冬季。在这10年间,人们提出了概率图理论、核方法、流形学习、稀疏学习、排序学习

等多种机器学习新方向。特别是在20世纪末和21世纪初,由于在搜索引擎、字符识别等应用领域取得的巨大进展,机器学习的影响力日益兴旺。其标志事件有:1997年Tom Mitchell机器学习经典教科书的出现,2010年和2011年连续两年图灵奖颁发给了机器学习的研究者Valiant和Pearl。

“三十年河东,三十年河西”。2006年以后,神经网络突破了3层网络结构限制,即所谓的深度学习,大幅提高了模型的表示能力,又恰逢大数据时代相伴而生的高计算能力,神经网络化身为深度学习,再次将分类能力提高到同时代其他模型无法匹敌的程度,有人将其称为人工智能的第3次春天。在机器学习的许多应用领域,深度学习甚至成为机器学习的代名词。虽然如此,时至今日,深度学习仍然只是机器学习的分支,无论其沉寂或者过热,都不可能逆转,而只能加速全部机器学习本身应用越来越普及,理论越来越深入的发展趋势。

2 深度学习的适应范围

理论上,神经网络深度越大,其表示能力越高,但是深度学习对于计算能力和训练数据的规模提出了极高的要求。2008年以前,计算机的计算能力和训练数据规模不具备大规模进行深度学习研究的条件。随着云计算、大数据的普及,具备了研究深度学习的外在技术条件。在2010年以后,人们通过采用新的激励函数,如ReLU,以及Dropout^[6],Batch Normalization^[7]等新训练方式,还有特别设计的新网络结构Deep Residual Networks^[8]等,逐渐克服了梯度消失或者发散问题,研究深度学习的内在技术条件也日渐成熟。这使得化名为深度学习的神经网络研究进入了另1个春天。

虽然如此,深度学习在理论上并没有突破以往神经网络的理论架

构。所有对于经典神经网络的理论分析对于深度学习也依然成立。1986年, Rumelhart 等人提出了自编码器, 该模型可以用来对高维数据进行降维^[9]。2006年, Hinton 等人在 Science 上发表了1篇文章, 该文章通过改进的自编码器学习算法构建了1种深层自编码器^[10], 自此深度学习的影响力日渐增大。常见的几种典型的深度学习网络包括: 自编码器、卷积神经网络、循环神经网络、长短时记忆网络等。

感知器算法可能是最早的神经网络算法, 该算法显然属于典型的白箱算法, 但是其表示能力有限, 连异或问题也解决不了。为了解决异或问题, 主流的神经网络技术放弃了解释性, 在黑箱算法的道路上越走越远。实际上, 机器学习算法对于普通人来说, 可粗分为两类: 一类是傻瓜型学习算法, 即只要输入一定, 任何人都可得到同样的结果, 如主成分分析等算法; 另一类是专家型学习算法, 即使输入相同, 不同人由于参数设置不同, 也会得到大不相同的结果。显然, 神经网络学习算法是典型的专家型学习算法。

总而言之, 机器学习有两个基本任务。一是试图发现输入和输出之间的因果关系, 其主要功用是解释, 最终目的是控制, 即一旦发生问题, 必须找出问题发生的原因, 这样就可以通过控制学习算法输入使得输出满足需要。解决此类任务的学习算法是白箱算法, 要求解释能力强。二是力图发现输入输出的相关关系, 其主要功用是预测, 最终目的是验证, 即一旦做出判断, 就可以根据外界反应判断预测是否准确, 但是出现错误之后, 并不要求根据输入来追踪错误发生的原因。解决此类任务的典型学习算法是黑箱算法, 并不需要解释能力。

真实现实生活中这两类任务都是存在的。第1类任务, 如各种高风险任务, 包括无人驾驶(火车、飞机、

汽车等)、医疗手术等, 一旦发生错误, 由于成本巨大, 必须能够分析出发生错误的原因, 以避免类似错误再次发生。完成这类任务, 不但需要提高完成任务的性能, 更重要的是能够发现输入与输出之间的因果关系, 一旦发生错误, 能追踪学习算法发生错误的原因, 显然适宜解决此类问题的学习算法是白箱算法。第2类任务, 如各种低风险甚至无风险性任务, 包括搜索引擎、各种棋牌游戏等, 显然这类任务即使发生错误, 后果也不严重, 成本可以承担, 因此更重要的是提高其性能, 特别是预测能力, 而并不要求算法去解释这些错误为什么会发生。

显然, 对于一个具体的学习任务, 一旦白箱算法的性能超过黑箱算法, 黑箱算法就再也不会是完成此类任务的优先考虑对象。但是, 许多学习任务, 由于具有极高的复杂性, 难以设计1个性能满足需要的白箱算法, 黑箱算法由于放弃了解释能力的约束而可能在性能上有较大优势。如今深度学习的表示能力已经十分强大, 2015年卷积神经网络已经达到152层^[8], 2016年卷积神经网络达到了1207层, 迄今为止没有任何一个白箱算法的表示能力可以与现今的深度学习相媲美。故可以预测, 深度学习在不需要发现因果关系的学习任务上在可见的未来不再有被替代的可能。

另外需要指出的是, 相关性的挖掘是目前大数据面临的典型任务。甚至有人认为, 在大数据时代, 数据相关性的重要程度远超数据因果性。由此可知, 相关性任务在大数据时代应用广泛。当前深度学习的快速发展和应用领域的日渐扩大, 从侧面证实了这一点。当然, 这并不意味着不需要研究数据因果性, 更不意味着数据因果性的消失。

3 结束语

深度学习不仅是目前热度最高

的人工智能研究方向, 也是工业应用最广泛的学习范式。在未来, 随着深度学习与特定相关性学习任务的耦合程度越来越高, 可以想像深度学习会有更多的变型出现。但是解释性的学习算法无论在工业界还是学术界同样也不会被放弃。

参考文献

- [1] MITCHELL T. Machine Learning[M]. New York: McGraw Hill, 1997
- [2] VAPNIK V N. The Nature of Statistical Learning Theory[M]. New York: Springer, 1995
- [3] HASTIE T, TIBSHIRINI R, FRIEDMAN J H. The Elements of Statistical Learning[M]. New York: Springer, 2003
- [4] THORNDIKE R L. Who Belongs in the Family[J]. Psychometrika, 1953, 18(4): 267-276
- [5] MINSKY M, PAPERT S. Perceptrons[M]. MA: The MIT Press, 1969
- [6] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors[J]. Computer Science, 2012, 3(4): 212-223
- [7] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]// Proceedings of the 32 nd International Conference on Machine Learning, 2015
- [8] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2016: 770-778. DOI: 10.1109/CVPR.2016.90
- [9] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning Internal Representations by Error Propagation[M]. Neurocomputing: Foundations of Research. MA: MIT Press, 1988: 318-362
- [10] HINTON G E, SALAHUTDINOV R R. Reducing the Dimensionality of the Data with Neural Networks[J]. Science, 2006, 313(9): 504-507. DOI: 10.1126/science.1127647

作者简介



于剑, 北京交通大学计算机学院教授, 交通数据分析与挖掘北京市重点实验室主任; 主要研究方向为机器学习、自然语言处理等; 曾著有《机器学习: 从公理到算法》一书, 已发表论文 100 余篇。

当深度学习遇到大视频数据

When Deep Learning Meets Big Video Data

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0044-003

摘要: 视频信号是大数据中的大数据, 这种海量视频数据带来了存储、传输、处理、管理等方面的挑战, 同时也提供了大量有价值的信息和商业机会。认为深度学习颠覆了视觉理解的进程, 从图像分类到物体检测、语义分割等更细更复杂的任务, 从视频里物体的检测与跟踪到物体属性和行为的分析, 特别是关于人和车的理解技术。指出随着计算能力和大数据持续快速增长, 加上深度学习、主动学习、迁移学习、无监督学习、强化学习等强大机器学习技术继续发展, 让机器可以像人一样看到并理解世界的前景是乐观的。

关键词: 深度学习; 大视频数据; 人工智能

Abstract: As the biggest big data, big video data imposes significant challenges to storage, transmission, processing, and management, but it also provides an enormous amount of rich information and huge business opportunities. Deep learning has fundamentally changed the way we understand visual information, from image classification to object detection and semantic segmentation, and from object detection and tracking to the analysis of object attributes and behaviors in video. This is especially true in understanding humans and vehicles. With the rapid growth of big compute, big data, and advanced machine learning techniques such as deep learning, active learning, transfer learning, unsupervised learning, and reinforcement learning, machines will soon be able to see and understand the physical world in the same way as humans do.

Keywords: deep learning; big video data; artificial intelligence

曾文军/ZENG Wenjun

罗翀/LUO Chong

(微软亚洲研究院, 北京 100080)

(Microsoft Research Asia, Beijing 100080, China)

- 视频数据已渗透到人类日常生活的方方面面, 视频分析的应用也因此是多方面的
- 深度学习颠覆了视觉数据理解的进程
- 基于深度学习的计算机视觉市场竞争日趋白热化。这种激烈竞争反过来将会进一步刺激加快计算机视觉和视频分析技术的发展

1 人工智能离不开视觉计算

人工智能(AI)是当今科技世界炙手可热的词语, 每个人都在谈论。在过去2~3年里, AI击败人类的新闻不断, 从Facebook的面部识别技术DeepFace达到和人类一样的识别精度^[1], 微软深度学习系统在图像识别上打败人类^[2], 机器在智商测试中击败人类^[3], 到AlphaGo击败围棋世界冠军李世石, AI的高热度在继续。

那么这些突破有哪些共性呢?

第一, 他们都归因于大数据的到来, 例如数千个小时有标注的语音数据, 数千万有标签的图像等; 第二, 离不开巨大的计算资源支撑, 包括图形处理器(GPU)和云集群的到来和普及。在此基础上机器学习技术才取得显著进展, 特别是深度学习的飞速发展。我们正处在AI的黄金时代。

AI离不开感知, 而视觉是我们最主要的感知手段。研究表明: 人的感知、学习、认知和活动有80%~85%是通过视觉介导的^[4]。如果不能获取并处理视觉信息, 就没法研究真实世界的人工智能, 由此可见计算机视觉对人工智能发展的重要性。

视频信号在大数据中占很大比重, 现在网络上70%~80%的流量是由视频信号所组成的, 可以说它是大数据中的大数据。这些数据可能在几年前还不太容易得到, 但是随着各种摄像头的普及, 视频数据得以更详细的记录物理世界发生的一切。由此产生了海量的大视频数据, 这种大数据给我们带来了存储、传输、处理、管理等方面的挑战, 同时也提供了很大的机会, 让机器帮助分析理解视频大数据就成了我们观察了解物理世界的一条捷径。现在我们通过分析这个大数据, 提取有价值的信息, 从而去支持新的产品或者服务, 所以这里

收稿时间: 2017-05-28
网络出版时间: 2017-07-05

面蕴藏了巨大的商业机会。视频数据已渗透到人类日常生活的方方面面,视频分析的应用是多方面的,包括居家、企业、零售、公共安全、交通、制造等,市场巨大。比如:预计全球家居安防解决方案市场将以8.7%的复合年增长率增长,到2020年将达到475亿美元^[5],半自动车市场预计到2018年将达到214亿美元^[6]。

2 深度学习颠覆了视觉理解的进程

视觉信号分析的发展起伏起伏,每到一定阶段都会出现“瓶颈”,其中很大的瓶颈就是没有足够量的数据,所以模型或算法的发展都受到一定的限制。直到大约2009年,ImageNet产生了。它是迄今为止最大的有标记的图像数据库,根据WordNet的层次结构组织,有超过10万个概念,每个概念有数百到数千幅附属的图像。ImageNet在过去几年大大促进了计算机视觉和图像分析的发展。

在ImageNet的基础上,近几年有一些与图像识别相关的挑战赛,如业界熟知的图像分类比赛就是利用100多万标注图像,进行1000种分类方法准确性比较的挑战赛。还有一些如物体检测、场景检测、场景分析和语义分割等基于ImageNet的比赛。

关于ImageNet图像分类比赛,在2012年前由于分类错误率很高,从而限制了它的实用。2012年,Hinton的实验室第1次把深度神经网络应用到图像分类任务上,其性能才得以大幅提升^[7],充分展示了深度神经网络对视觉研究的极大潜能,也掀起了视觉研究的新高潮,让人们看到了计算机视觉实用化的希望。短短几年后的今天,深度神经网络技术发展迅猛,在ImageNet图像分类上的性能已超过人类,人们研究的重点也从图像分类转移到图像物体检测、语义分割等更细、更复杂的任务。

图像分析已经有了很大的进步,视频分析和理解方面进展则稍显缓

慢。视频信号相比于图像信号有更大的挑战,因为它是一个更高维的信号,所含内容的多样性也很复杂,所以要去判断它、理解它都很困难,当然数据量很大也是另外一个原因。除此之外,在很多情况下视频是提供实时监测控制的,因此对处理速度等指标也有很高的要求,加之标注视频数据时每1帧都要标注,费时、费力且成本高昂,视频发展相比图像来讲还是落后一些。当然,如何获得足够多训练数据也是必须解决的难点。

前面谈到视频分析的一些应用场景,尽管不同应用场景有不同技术要求,但有些基本技术是共享的,比如物体的检测与跟踪。人是我们的日常生活和工作的核心,因此也是大多数图像/视频的最主要实体。对人的分析是视频理解中的关键一步。因此很多研究团队包括微软亚洲研究院最近几年都专注于以人为中心的视频分析,例如检测与识别人、人的属性、人的行为,甚至是人的意向。由于近年来大数据、计算能力和深度学习技术的进步,对视觉数据中人的理解技术已取得了很大的进步。机器检测和识别人脸的性能已经达到了和人相仿的程度,并在身份验证、安全、智能零售、智能媒体管理等领域得到广泛应用。人体检测的性能也有了显著提高,在一些基准数据集上达到超过80%的准确度。人的各种属性,如性别、年龄、情感、手势与身体姿势,以及衣服颜色类别等也可以很好地提取,以帮助更好地了解一个人的状态。人体姿态估计技术的性能也达到了数年前不可想象的水平,并极大地方便了人的动作识别。

3 车辆和车牌检测与识别案例

日常生活中,尤其是城市生活中,车辆是重要性仅次于人的目标类别。深度学习技术的飞速发展也大大的带动了与车辆相关的计算机视觉技术的发展,其中,最重要的就是

车牌和车辆的图像检测与识别技术。

车牌是车辆的身份证,车牌自动识别技术有着非常广泛的应用,例如:车辆进入管控区域时的权限验证,进入停车场或高速公路时的收费管理,或者道路车辆违章摄像。目标通常分为合作目标和非合作目标。合作目标的图像检测和识别技术已经相当成熟,在大量应用的车牌识别系统中包含图像采集、车牌检测、字符抽取和字符识别4个步骤,其中图像采集环节是可控的,即图像采集对象是合作目标。比如车辆进入停车场时,需要车辆在低速甚至完全静止状态下完成图像采集,而且拍照时车牌的位置相对固定,这就在最大程度上保证了图像清晰,同时有效限制了车牌检测时的搜索范围。

然而,在大数据时代出现了一些车牌检测的新应用需求。这些应用面对的是不可控的图像采集对象,即非合作目标。图像质量良莠不齐,车牌类别、大小、出现位置、光照条件等都有很大的不确定性,如图1所示。例如:交通管理部门希望能够从公交车摄像头获取的图像数据中自动提取违章占用公交专用道车辆的号牌信息,然而公交车摄像头获取的是非合作目标的图像,基于合作目标的图像检测和识别技术显然无法满足应用需求。利用我们在车牌检测技术上最新的基于深度学习的研究成果,可以准确、高效的解决这一难题。

图2展示了一些我们获得的车牌检测结果示例。我们的方案可以在不同的光照条件下准确定位到大小、视角不同的各类车牌的4个角点。

视频车牌模糊也是一个非常典型的新应用需求。用户在视频网站分享视频时,如果镜头中有车辆出现,用户希望能够模糊掉车牌信息以免侵犯他人隐私,就需要利用视频车牌模糊技术,其中车牌检测和跟踪是关键所在。基于深度学习的图像车牌检测和跟踪可大大提高视频中车牌的召回率,提升车牌模糊的性能。



▲ 图1 不可控的图像采集对象



▲ 图2
典型的车牌检测结果

图像和视频中的车辆检测因其在自动驾驶、道路监测控制中的应用受到了广泛的关注。然而,深度神经网络的出现使车辆检测的精度有了质的提升。KITTI 是车辆检测领域一个著名的公开数据集。在深度神经网络被大规模应用到物体检测领域之前,Regionlets^[8]曾作为一个标杆方案,获得了较高的检测准确率。其在简单、中等难度和较难数据类别上的准确率分别为 86.5%, 76.56% 和 59.82%。然而近年来,随着 Faster R-CNN^[9]模型的提出,Regionlets 在 KITTI 车辆检测排行榜上已退居到第 50 名的位置。截至目前,在中等难度的车辆检测上已有超过 10 种方案可以获得超过 90% 的准确率。另外,深度神经网络的出现也推动了车型车款识别(定位到车型车款),车辆精细化识别(定位到具体车辆)等方向的发展,使得智能城市的构想不再遥远。

总之,随着计算能力的持续快速增长,加上深度学习、主动学习、迁移学习、强化学习等强大机器学习技术继续发展,让机器可以像人一样看到并理解世界的前景是乐观的。

4 结束语

再好的研究成果,最终只有在实

际应用中得到验证才能体现它的真正价值。微软亚洲研究院研发的视频分析技术正在通过微软认知服务这个平台,以视频应用程序编程接口(API)的形式提供给广大人工智能领域的开发者,帮助大家方便而高效地开发和视频相关的人工智能应用系统。这些技术也已成为微软 Azure 云平台的媒体分析服务的重要组成部分,可提供企业级的智能服务。类似的,其它高科技公司如 Google、Amazon、Facebook 等也相继推出基于深度学习的计算机视觉 API,从而使使得市场争夺日趋白热化。这种激烈竞争反过来将会进一步刺激加快计算机视觉和视频分析技术的发展,最终使人工智能更快、更深入地渗透到人类日常生活和工作中去。

参考文献

- [1] Sophos. Facebook's DeepFace facial recognition technology has human-like accuracy[EB/OL]. (2015-02-06)[2017-06-11]. <https://nakedsecurity.sophos.com/2015/02/06/facebook-deepface-facial-recognition-technology-has-human-like-accuracy/>
- [2] NOVET J. Microsoft Researchers Say Their Newest Deep Learning System Beats Humans — and Google[EB/OL]. (2015-02-09)[2017-06-11]. <https://venturebeat.com/2015/02/09/microsoft-researchers-say-their-newest-deep-learning-system-beats-humans-and-google/>
- [3] MIT Technology Review. Deep Learning

Machine Beats Humans in IQ Test[EB/OL]. (2015-06-12)[2017-06-11]. <https://www.technologyreview.com/s/538431/deep-learning-machine-beats-humans-in-iq-test/>

- [4] Brainline. Vision Problems[EB/OL]. (2017-06-11). http://www.brainline.org/landing_pages/categories/vision.html
- [5] Markets and markets. Home Security Solutions Market – Global Forecast to 2020 [EB/OL]. (2017-03)[2017-06-11]. <http://www.marketsandmarkets.com/Market-Reports/home-security-solutions-market-701.html>
- [6] Markets and markets. Semi Autonomous Market for Passenger Car-Global Trends & Forecast to 2018[EB/OL]. (2017-05)[2017-06-11]. <http://www.marketsandmarkets.com/Market-Reports/near-autonomous-passenger-car-market-1220.html>
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105
- [8] WANG X, YANG M, ZHU S, et al. Regionlets for Generic Object Detection[C]// IEEE International Conference on Computer Vision. USA: IEEE Computer Society, 2013: 17-24. DOI: 10.1109/ICCV.2013.10
- [9] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149. DOI: 10.1109/TPAMI.2016.2577031

作者简介



曾文军,微软亚洲研究院副院长,IEEE Fellow,密苏里大学、中国科技大学、西安交大、天津大学等多所学校博士生导师,2003—2016 任密苏里大学计算机科学终身教授;目前负责微软亚洲研究院视频分析和理解的研发,为微软认知服务和 Azure 媒体分析服务提供技术;曾对国际标准 ISO MPEG、JPEG2000 和 Open Mobile Alliance 的发展作出重大贡献,担任多个 IEEE 期刊和杂志的副主编、多个 IEEE 会议共同主席或技术程序委员会主席;已发表 160 篇论文,并有 2 部关于多媒体安全和社交多媒体的著作。

罗静,微软亚洲研究院网络多媒体组主管研究员,中国科学技术大学兼职教授、博士生导师,IEEE 高级会员;主要研究方向为多媒体云计算、多媒体通信、计算机视觉等;自 2011 年起担任 IEEE Infocom 会议的技术委员会成员;已在 ACM Mobicom、IEEE Infocom 等顶尖学术会议以及多份 IEEE 期刊上发表论文 30 余篇,拥有 10 余项国际发明专利。



罗静,微软亚洲研究院网络多媒体组主管研究员,中国科学技术大学兼职教授、博士生导师,IEEE 高级会员;主要研究方向为多媒体云计算、多媒体通信、计算机视觉等;自 2011 年起担任 IEEE Infocom 会议的技术委员会成员;已在 ACM Mobicom、IEEE Infocom 等顶尖学术会议以及多份 IEEE 期刊上发表论文 30 余篇,拥有 10 余项国际发明专利。

关于人机对话系统的思考

Thoughts on Human-Computer Dialogue System

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0047-004

摘要: 提出了一系列非常重要、影响人机对话质量的问题,包括:如何面向自然语言理解(NLU)构建对话任务分析、深度推理,如何利用语言学尤其是互动语言学研究成果构建对话管理(DM),如何有效建模人机对话中不同任务间的关约束来发展联合模型等。认为尽管人机对话系统的基础模型已取得了长足进步,但如果不能有效地解决上述问题,就不可能获得高质量的人机对话系统,自然语言处理的水平也就难以得到实质性提升。

关键词: 人机对话系统; NLU; DM; 自然语言生成(NLG)

Abstract: In this paper, a series important problems which have significant influence on the quality of human-computer dialogue are proposed, including how to normalize task analysis and introduce deep reasoning for natural language understanding (NLU), how to build linguistically, especially interact linguistically reasonable dialogue management (DM), and how to develop joint method for modeling correlations between different subtasks in dialogues. There has been lots of progresses on human-computer dialogue, but if the problems mentioned above cannot be solved, it is difficult to achieve high quality dialogue systems and important improvements in natural language processing.

Key words: human-computer dialogue system; NLU; DM; natural language generative(NLG)

王小捷/WANG Xiaojie

(北京邮电大学, 北京 100876)
(Beijing University of Posts and
Telecommunications, Beijing 100876,
China)

- 对话任务分析的自动化、深度推理的缺乏是自然语言理解面对的重要挑战,目前相关工作较少,需要大力发展
- 对话管理应汲取会话分析、互动语言学等方面的研究成果,研究相关计算模型,才有可能构建自然和谐的人机对话系统
- 联合建模人机对话系统多个子任务的方法比传统的管道方法具有诸多的优势,深度学习提供了有效联合建模工具,已成为一个重要的技术发展趋势

人类社会的一个主要特征是人际交流,广泛深入的人际交流是促进合作,推动社会向前发展的重要动力。人际交流的主要工具是包括汉语、英语在内的各种自然语言,基于自然语言的对话是人际交流的主要方式之一。基于自然语言的对话也称为会话,可以是在两人之间进行,也可以在多人之间进行。此外,自说自话的独白也常被语言学家归入会话。

近年来,随着技术日新月异的发展,各种人造机器成为人类的亲密伙

伴,人与机器的交流变得越来越重要。为了与机器交流,让机器明白人类需要它们做什么,人类发明了很多种机器语言。开始,这些机器语言只是少数人使用,随着机器进入到普通人的日常社会生活,就需要机器能与普通人进行交流,发明越来越简单的机器语言或操作方式是一个途径,而让机器能懂得人类语言无疑是一个更自然、更令人向往的方式。人机对话系统旨在实现这种方式,满足普通人与机器进行顺畅交互的需求。

人机对话系统经历了 50 多年的发展,取得了不少的进展,目前更是面临巨大的发展机遇。如何抓住机遇,促成人机对话系统的进一步发

展,是每位研究者都需要认真思考的问题。

1 人机对话系统简述

20 世纪 60 年代,Weizenbaum 创建了 ELIZA 对话系统^[1],试图模仿心理治疗师与人进行对话交流;90 年代,Senef 等人开发了自动机票预订系统(ATIS)^[2],通过与用户进行多轮对话来获取用户需求(如出发地、出发日期、航班号等),以帮助用户预订机票。这两个系统分别代表了两类不同的对话系统,后者属于目标驱动型对话,前者属于非目标驱动型对话。

目标驱动型人机对话系统通过与用户进行多轮基于自然语言的对话

收稿时间: 2017-05-26
网络出版时间: 2017-07-05

话来逐步收集与目标相关的信息,辅助用户成功获得某种服务。与 ATIS 系统类似的还有电信服务系统 HMIHY^[3]、天气信息查询系统 JUPITER^[4]、旅行计划制定系统 DARPA Communicator^[5]等。目标驱动型的人机对话系统也称为面向任务的人机对话系统。

而以 ELIZA 系统为代表的非目标驱动型人机对话系统通常只是对用户输入进行基于自然语言的响应,而不完成特定信息服务任务,类似于闲聊似的对话,所以也称为聊天机器人。与目标驱动型系统相比,这类系统对收集信息、自主确定对话行为等方面的要求都较弱。

一个典型的任务驱动型人机对话系统的技术模块如图 1 所示^[6],系统主要包含自然语言理解(NLU)、对话管理(DM)、自然语言生成(NLG)3个部分。用户输入可以是语音,也可以是文本,如果是语音,则 NLU 部分包含语音识别模块;而系统输出为语音时,NLG 部分还需包含语音合成模块。NLU 部分进一步又分解为3个子任务:其一是领域识别,判定用户谈论的是什么领域的事情;其二是意图识别,识别用户话语的目的,比如是告知某个信息,还是确认某个信息;其三是语义分析,又称为槽填充,旨在标识用户话语中与目标有关的语义类别,比如预订机票时需要的出发地、目的地、时间等语义类别。DM 则进一步分解为对话状态追踪和动作选择两个子任务;对话状态追踪获取对话的当前状态;动作选择确定系统响应的动作。DM 模块负责控制整

个对话过程,例如:决定系统在上一轮对话中获取了用户出发地信息之后,下一步该和用户说什么。NLG 模块则把决定要说什么具体转化为自然语言,通常也进一步分解为几个子任务。

非目标驱动的对话系统大致按此结构,但是大多没有显式的意图识别、语义标注等 NLU 过程,也没有显式的 DM。

在实际应用中,目标驱动与非目标驱动的划分也并不是绝对的,常常会需要目标驱动和非目标驱动混合出现。比如:目前的个人助理系统 Apple Siri、Microsoft Cortana 等,有时可能要与用户闲聊,这时是非目标驱动的,有时可能要帮助用户订票,这时是目标驱动的。这一系统框架经历了一系列的发展,其中的各个技术部分也在飞速发展^[7-9]。

2 自然语言理解

人机对话系统中 NLU 的主要目标是识别输入话语的领域和意图,获得任务相关的语义信息。例如:在订机票时,客户说“订1张5月10号去北京的票”,其领域是“航班信息服务”,意图是“订票”,与订票任务相关的语义信息包括作为“订票数量”的“1张”、作为“起飞时间”的“5月10号”,以及作为“目的地”的“北京”。订票数量、起飞时间和目的地称为槽,句子中提供的这些槽的具体值称为槽值。

如何准确地进行领域、意图和槽值的识别或标注是当前针对人机对话的 NLU 研究中得到关注最多的方

面。近年来人们不断设计新的模型和算法来提升语言理解的性能^[9],而另一个研究重点是当面对的对话任务发生变化时,如何快速准确地将模型推广到新的对话领域中,即模型的跨领域适应问题。

然而,笔者认为,目前的语言理解中还存在1个问题没有得到应有的重视,即:如何为 NLU 模块设定目标。从前面的例子可以看到,目前的语言理解对于给定的用户输入通过分析获得什么信息都是基于对话任务而确定的。比如:NLU 算法只会在句子中去标注订票数量、起飞时间和目的地等槽的槽值,而不会去标注“运动类型”等无关的信息。但是对话系统如何能知道它只应该去关注这些信息,而不应关注其他信息呢?目前的方法是完全由人预先基于对话任务来确定语言理解的目标,再基于这些目标来训练模型。这种方法需要预先对任务有完整的分析,建立关于任务的完整意图和槽的体系。这类似于软件工程中在构建一个软件时需要完整的需求分析一样。对于软件工程,有一套需求分析技术,而对于人机对话系统,目前还并没有这样的任务分析技术,一般的人机对话系统的任务分析还停留在经验阶段,这给人机对话系统的普及带来了困难。

进一步,在目前的 NLU 中,对于语言的理解还停留在比较浅的层次。例如:在槽标注时,大多数模型还只能在语言表层进行标注,在复杂情况下结合推理进行深入理解的能力还较弱或者基本没有。例如,客户说“订一张端午节后一天的票,如果那天预报有大雾的话就往后推一天。”这时简单地在语言表层标注时间槽已经不太可能,需要结合其他信息并通过推理得到正确的槽值。语言推理不仅需要对话语言本体的理解,其背后还需要更为丰富的关于现实世界对象、事件及其关系的知识进行支撑,并且有的知识是显式的,而有

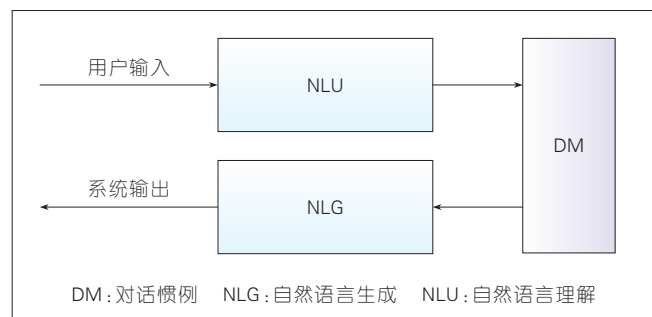


图1
人机对话系统框架

些知识是隐式的。例如,上一句中的“端午节”“下雨”等是现实世界中的对象和事件,其中,端午节具体是哪一天可以直接查询获得。但是其中还含有简介的、隐式的推理知识,比如在上句中人可能会依据如下的隐含逻辑去理解说话人:如果有大雾的话,飞机可能不能按时起飞,客户大概不希望延误,所以决定在这种情况下推迟一天,推迟的目的是为了等一个好天气。如果知道存在这样一个隐含的推理,那么作为好的客服人员,可能就要再问客户:如果后一天还报有大雾怎么办?这类知识以及要基于这类知识进行的推理将导致对语言的理解更为复杂和困难。

3 对话管理

DM 追踪对话状态,并基于此决定系统的响应动作。对话状态追踪和动作选择的研究已得到了广泛关注,取得了大量进展。这些研究一方面由 DM 计算模型的发展所带来^[79],例如,从马尔科夫决策过程(MDP)到部分可观马尔科夫决策过程(POMDP),从 Q-Learning 算法到深度 Q 网络(DQN);另一方面也离不开对语言学研究的有效借鉴,毕竟人机对话是希望建模人人对话,只有对人类对话的本质有了更深刻的理解,才有可能在人机对话的研究中取得相应的进展。这方面,已有的借鉴包括会话分析研究中话轮、相邻对等概念,它们已成为目前计算模型的建模单元,此外,言语动作等语用学的概念也已成为 DM 的决策目标。

但是,更多、更深入的语言学研究成果还没能在计算模型中得到有效体现。如格莱斯的对话合作原则、礼貌原则等还很少进入计算模型的考量范围。近 20 年来,一些新兴的关于语言的理论也还几乎没有进入计算语言学家的视野,其中就包括笔者认为对于对话建模和 DM 具有重要借鉴价值的互动语言学。

互动语言学^[80]考察和研究日常真

实人类互动过程中的语言使用情况,探讨语言结构及其运用模式与互动交流是如何相互影响的,为互动参与者在真实、情景化的日常交往中完成互动与交流提供理论解释和支持。互动语言学者从功能语言学、会话分析的成果以及对真实对话数据的研究出发,对传统的语言学有直接的批评。传统的以乔姆斯基生成语言学为代表的语言学者认为以语言能力为基础的语言结构的理想形式是语言学研究的对象,语言行为是语言能力在具体使用情景中的实现。互动语言学者则认为,语言的结构和意义是在情景互动过程中产生,并随互动的发展不断变化,是说话者和接受者共同协商的产物。例如:形式语言学将句子这一语言单位作为单个说话者语言能力的产物;而互动语言学者则认为,句子的生成是说话者和接受者在交际过程中互动的产物,它不仅与说话者有关,还关系到接受者,因为说话者在互动中往往需要根据接受者的反应调整话语^[81]。

这种语言结构和意义可以基于互动而不断进行建构的观点如果能用于 DM,帮助构建更好的对话结构,进行更好的对话言语行为决策,将是十分有价值的。至于如何借鉴,则还需要深入了解互动语言学的相关成果,同时综合考虑在当前科学与技术发展水平下,构建其数学模型和计算模型的可行性。

笔者在此仅举一个可能的例子:交互语言学认为语言结构是在语言交互中通过互动而产生的,那么一个好的对话行为决策所基于的对话语言分析和理解可能不应仅局限在一个说话人话轮中进行。在对话中,将对话双方的邻接话轮,即相邻对,作为一个分析单元可能更为可取。目前的模型是可以基于多轮对话信息来进行对话行为选择。但是,通常只是简单地拼接对话历史,而没有考虑对话历史中的结构信息。而如果以相邻对为基础单元,在计算模型上可

以构建关于对话的层次结构,有利于建模对话的语言结构,从而为对话行为提供更准确的决策信息。

4 系统构建

传统上,一般采用分而治之的方式实现如图 1 所示的人机对话系统。即:先分别构建 NLU、DM 和 NLG 等模块,这些子模块通常还要分解为更小的子任务分别建模,之后按序将这些模块连接起来。这种方案的优势是简洁清楚,各部分的任务明确,可以分别开展研究,各自解决各自的问题。

但是这种研究方案的问题也很明显。首先,各个子任务分别研究导致难以建模子任务之间的关联约束信息,而有时这种关联约束信息是非常有价值的。例如:在 NLU 模块中的意图识别和语义分析两个子任务就是紧密相关的,一个具有告知所需航班起飞时间意图的句子很可能包含时间信息,反之,包含时间信息的句子很可能是表明了一种告知时间的意图。这两个子任务的交互约束信息是可以互相帮助,共同提高彼此性能的。而如果分别建模,这种交互约束信息就不能得到充分利用了。

其次,各部分分别研究时,某个任务产生的错误结果只能用来调整这个任务本身,而不能有效地反传到前面的子任务上,而实际上后一任务的错误很可能是前面子任务错误的累计导致的。例如:在 DM 的决策出现错误时,其原因可能是状态追踪发生错误,也有可能是语言理解部分发生了错误,如果分别建模,则 DM 的错误不可能直接用于对 NLU 部分的错误进行修正。

因此,人机对话系统的一个重要发展方向是多任务的联合建模。联合可以有不同的层次,例如:NLU 模块内各个子任务的联合建模,综合利用意图识别和语义分析两个子任务之间的交互约束信息,可以获得更好的理解性能;综合 NLU 与 DM 的跨模

块联合建模,使得DM的决策错误不仅可以用于修正状态追踪的错误,也可以直接用于修正语言理解的错误。

另一方面,近年来深度神经网络研究的发展也为联合建模提供了重要的技术工具。2015年出现了多种基于序列到序列深层神经网络模型的非目标驱动型对话系统。图2是一个简单的例子,其中输入句子在编码后逐步解码得到机器输出的响应。初期的模型大多只是处理一问一答的应答系统,而随着输入中加入对话历史,应答系统逐步向对话系统发展。与此同时,随着当前深度强化学习的发展,面向目标驱动型对话系统也逐渐成为可能。

这种跨越NLU和DM的联合建模具有特别重要的意义,因为它将语言输入和决策(说什么)直接关联起来了,跨越了NLU任务。实际上,我们在人际对话中是难以看到人的语言理解结果是什么,而只能看到决策的结果是什么。我们目前定义的NLU任务只是我们给人机对话引擎设计的,人在对话中是否有类似的理解过程和结果其实是不清楚的,也有可能不存在或与我们定义的不一样。我们并不是要否定NLU任务定义本身,而是认为,如果未来能发展出直接关联语言输入和决策结果的好模型,那么基于对这种模型内部状态的分析,也许可以帮助我们了解在对话过程中真正需要的NLU任务是什么。

5 结束语

人机对话系统业已展现出广泛的应用价值,也在很多实际问题中发挥了作用。在微软Build 2016开发者大会上,微软首席执行官纳德拉提出了作为平台的对话(CaaP),将对话作为平台技术,作为整合人工智能技术

的途径,这吸引了非常广泛的研究和商业关注。

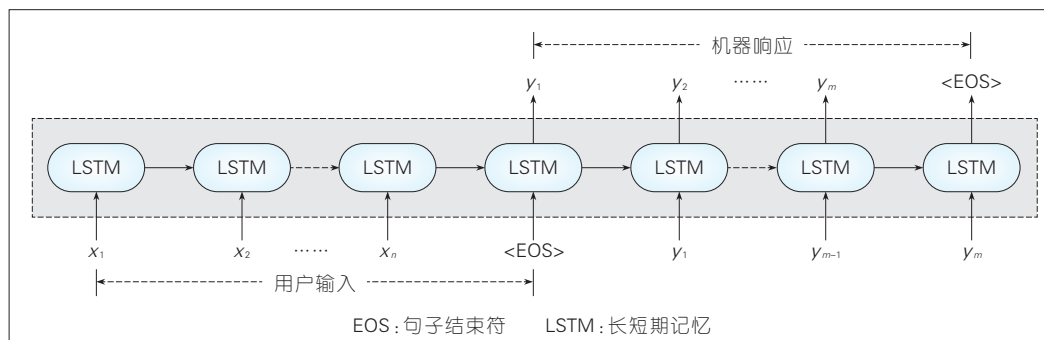
从目前的问题研究和技术水平来看,为特定任务提供信息服务的人机对话系统,在进行精心的任务定义和设计之后,是可以获得较好的服务质量的,但是其鲁棒性不能得到有效保证。处于任务边界的话语将很可能导致系统不能优雅地退出工作,而对于更为复杂的多任务、跨任务等情形,系统的鲁棒性就更难以保证了。然而,反观人类的日常人际对话,任务边界模糊在大多数情况下并不是问题,除非进入非常专业的领域,跨领域并不太影响人类的人际交互能力,鲁棒性似乎是人际对话的一个基本特质。人际交互时不懂可以马上就问,这本身也就是维持对话的一种方式,而不像在阅读时遇到不懂问题就难以继续推进阅读本身。这些迹象表明,也许目前我们还远没有触及到人机对话系统的核心问题,未来还需要更多、更深入的工作。

参考文献

- [1] WEIZENBAUM J. ELIZA—A Computer Program for the Study of Natural Language Communication Between Man And Machine [J]. Communications of the ACM, 1966, 9(1): 36–45. DOI:10.1145/365153.365168
- [2] SENEFF S, HIRSCHMAN L, ZUE V W. Interactive Problem Solving and Dialogue in the ATIS Domain[C]// The Workshop on Speech and Natural Language. Association for Computational Linguistics, 1991:354–359. DOI:10.3115/112405.112743
- [3] GORIN A L, RICCARDI G, WRIGHT J H. How May I Help You?[J]. Speech Communication,

1997, 23(1–2):113–127. DOI:10.1016/S0167-6393(97)00040-X

- [4] ZUE V, SENEFF S, GLASS J R, et al. JUPITER: A Telephone-Based Conversational Interface for Weather Information[J]. Speech & Audio Processing IEEE Transactions on, 2000, 8(1):85–96. DOI: 10.1109/89.817460
- [5] WALKER M, ABERDEEN J, BOL J, et al. DARPA Communicator Dialog Travel Planning Systems: The June 2000 Data Collection[J]. Veterinary Pathology, 2001, 9(3):190–201
- [6] JURAFSKY, D. MARTIN, J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition[M]. New Jersey: Prentice-Hall
- [7] YOUNG S, GASIC M, THOMSON B, et al. POMDP-Based Statistical Spoken Dialog Systems: A Review[J]. Proceedings of the IEEE, 2013, 101(5):1160–1179. DOI: 10.1109/JPROC.2012.2225812
- [8] YU K, CHEN L, CHEN B, et al. Cognitive Technology in Task-Oriented Dialogue Systems: Concepts, Advances and Future[J]. Chinese Journal of Computers, 2014, 37(18): 1–17
- [9] WANG X, YUAN C. Recent Advances on Human-Computer Dialogue[J]. CaaI Transactions on Intelligence Technology, 2016, 1(4):303–312. DOI: 10.1016/j.trit.2016.12.004
- [10] 林大津, 谢朝群. 互动语言学的发展历程及其前景[J]. 现代外语, 2003, 26(4):411–418
- [11] 罗桂花. 互动语言学:语言产生于互动 互动塑造语言[N]. 中国社会科学报, 2012–10–8 (A07)



▲图2 序列到序列的应答系统

作者简介



王小捷, 北京邮电大学教授、博士生导师、智能科学与技术中心主任, 中国人工智能学会常务理事、自然语言理解专业委员会主任; 主要从事自然语言处理、人机对话、多模态认知计算等方面的研究; 已在相关领域发表学术论文200余篇。

移动通信 Small Cell 基站供电解决方案与趋势分析

Power Supply Solutions and Trends Analysis for Small Cell Mobile Communication Base Station

胡先红/HU Xianhong
刘明明/LIU Mingming

(中兴通讯股份有限公司, 广东 深圳 518057)
(ZTE Corporation, Shenzhen, 518057, China)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0051-005

摘要: 小型分布式基站 Small Cell 数量的快速上升对供电方案提出了“零占地”、易部署等新的需求。针对不同的 Small Cell 基站应用场景, 提出了相应的供电解决方案, 包括本地交流(AC)供电、本地直流(DC) 48 V 供电、本地 AC 和 DC 48 V 混合供电、DC 拉远供电、以太网供电(POE)供电等。此外, 还指出了 Small Cell 基站供电方案的发展趋势, 包括: 小型化、模块化、高效节能、绿色、共享、智能互联等。

关键词: Small Cell; 供电解决方案; 零占地; 模块化

Abstract: With the rapid growth in number of small cell, new requirements are proposed, such as zero occupation of land and easy deployment. For different application scenes, different power supply solutions are proposed, including local alternating current(AC) power supply, local direct current(DC) 48 V power supply, local AC and DC 48 V mixed power supply, DC remote feeding power supply, power over ethernet(POE) power supply, and so on. Moreover, the trends of power supply solutions for small cell are pointed out, including miniaturization, modularized, high efficiency and energy conservation, green, sharing, intelligent and interconnection.

Keywords: Small Cell; power supply solution; zero occupation of land; modularized

随着移动通信网络的发展, 移动通信基站将围绕人们居住、工作、休闲、交通以及垂直行业的需求, 广泛部署在密集住宅区、办公室、体育场、地铁、高速公路、高速铁路以及环境监测等场景^[1]。移动通信分布式基站的形态、功耗、站址、覆盖等都在不断地变化。4G 比 3G 在带宽和通信速率有巨大的提升, 而 5G 将比 4G 具有更高的带宽、更低的时延、更多的连接数, 5G 基站数量将是 4G 的数十倍, 这些特征给供电系统带来新的要求。为适应这些特征, 需要为新型基站提供合适、经济的供电解决方案。

1 移动通信基站形态的变化

为了实现定向覆盖、增强覆盖、盲区覆盖、边缘覆盖的高密度、大流量通信业务, 小型分布式基站 Small Cell 的数量呈快速上升趋势。2012 年底, Small Cell 基站部署数量超过宏基站, 2016 年部署的基站中, 约 88%

是 Small Cell 基站, 大约 9 100 万个, 比 2015 年增长约 50%。由于数据业务的快速增长, 预计未来 Small Cell 基站还会快速增长。

Small Cell 基站指从产品形态、发射功率、覆盖范围等方面都比传统宏站小得多的基站类型。经统计分析, 绝大多数的数据业务发生在室内或热点区域^[2]。小基站可有效改善深度覆盖, 增加网络容量, 提升用户体验, 因而越来越多地受到业界的关注。Small Cell 基站系列产品包括 Femto Cell、Pico Cell、Micro Cell、Metro Cell 系列产品等, 各类站点的发射功率、典

型功耗等指标见表 1。

2 Small Cell 基站供电的主要需求

Small Cell 基站数量巨大, 应用场景多样, 与之配套的供电系统需要适应这些特征, 并满足以下主要需求。

2.1 “零”占地

Small Cell 基站广泛部署在商业区、办公区、住宅区等, 这些区域普遍存在业主协调难、站址资源不足的问题, 部分站点还需要考虑隐蔽的需求。因此, 需要 Small Cell 基站及供

收稿时间: 2017-05-27
网络出版时间: 2017-07-07

▼表 1 Small Cell 基站的分类

分类	发射功率	覆盖半径	用户容量	典型功耗	应用场景
Femto Cell	10~100 mW	10~20 m	4~8	15 W	室内小面积覆盖,如家庭、SOHO等
Pico Cell	室内型: < 250 mW; 室外型: < 1 W	室内型: 30~50 m; 室外型: 50~100 m	32~64	30 W	室内或室外中小面积覆盖,如中小企业、商铺、公共热点等
Micro Cell	5~10 W	100~300 m	< 100	100 W	室外补盲覆盖,如密集城区宏覆盖阴影区域、小型建筑空分
Metro Cell	5~10 W	100~300 m	100~200	150 W	室外补热覆盖,如密集城区街道、大型体育场等

电系统尽量少的占地或者不占地,降低站址选择租用困难以及减少站址使用费。例如:在城市街道,利用现有的灯杆进行抱杆安装,就可达到零占地的需求。

2.2 “零”维护

Small Cell 站点数量庞大,分布范围广,仅维护上站的交通费用就非常巨大。因此,需要供电系统尽量少维护,或者非常方便维护,甚至免维护,以降低运维成本。同时,在模块损坏时,可以非常方便更换。在多模块并联应用时,单模块的损坏不影响其它模块的工作,且更换时负载不断电。例如:供电系统采用自然冷却设计以减少风扇等易损部件的维护成本,模块损坏 10 min 内即可完成单站更换维护。

2.3 易部署

Small Cell 站点部署环境复杂,有街道、弱电井、高楼房屋外墙、室内天花板等,这些场景无法使用大型施工工具。因此,站点安装需要非常简单易行,工作量小,单人可操作,基本不需复杂调试,且一次上站就能完成安装部署。对于壁挂或抱杆安装,要求设备轻,单人单手可提举。普通资质工程安装施工人员经简单培训即可操作。

2.4 易扩展

Small Cell 站点的部署存在需要根据用户容量进行扩容的需求,或者先期部署 4G,后期部署 5G 共站的情

况。这就需要根据业务发展需要,通过并联,扩展供电容量,以满足设备分批部署、后期更多的设备扩容需求,从而可以在初期进行低配置,后期根据业务发展逐步扩容,减少初期投资浪费。例如:包括电源模块、电池组等在内的供电设备都需要模块化设计,可多模块并联扩容,且现场并联易操作。

2.5 云管理

广泛分布的 Small Cell 站点,处于无人值守状态,需要对供电系统的设备状态(市电是否正常、设备是否正常等)、供电参数(电压、电流、功率等)、环境参数(温度、湿度等)等实现远程监测控制,相关信息上传云端,实现云化管理。为降低通信成本,信息传输通过 Small Cell 通信设备实现,即带内传输。

3 Small Cell 基站供电解决方案

为适应不同 Small Cell 基站的供电需求,以及部署场景的不同,需要多样化的供电解决方案以满足 Small Cell 基站部署的要求。

3.1 本地交流(AC)供电解决方案

对于 AC 供电的 Small Cell 站点设备,在站点能够方便获取市电的情况下,有两种方案。一是不需要备电的情况下,可直接采用市电接入供电,这是最为经济的。市电停电会导致业务中断,通信质量降低,此方案在对市电条件好且通信质量要求不高

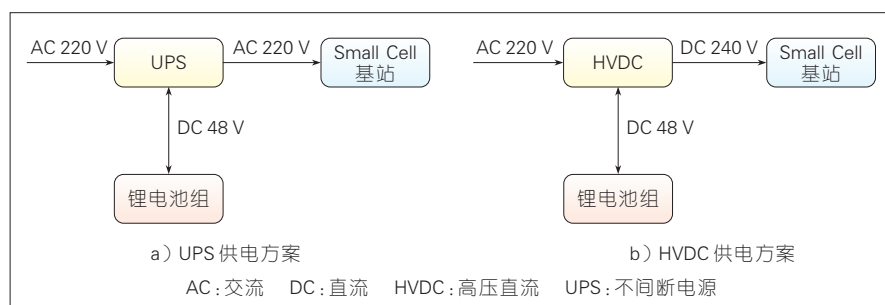
的通信站点可以采用。二是在需要备电的情况下,可采用交流不间断电源(UPS)供电,如图 1 a)所示。此外,当初次部署不需要备电,后期随着站点的业务发展,重要性提高,改造为有备电的情况,也可采用此方案。此外,当前大部分支持 AC 输入的 Small Cell 基站设备一般兼容 240 V 高压直流(HVDC)输入,也可以采用 HVDC 电源(带电池)提高通信基站供电可靠性,如图 1 b)所示。考虑电池成本和通用性,宜采用 48 V 电压等级的电池组作为备电,这种情况下电源设备的内部复杂度会有所提高。

3.2 本地直流(DC)48 V 供电解决方案

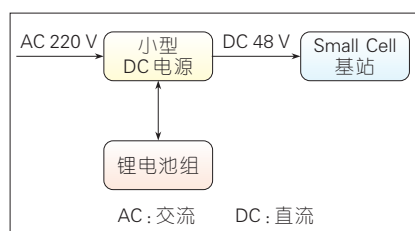
对于 DC 48 V 供电的 Small Cell 基站通信设备,在站点能够方便获取市电的情况下,可采用小型 DC 电源供电,根据站点市电的可用度、用户对业务可用度要求,及综合投资成本,选用配备或者不配备蓄电池组的方案,如图 2 所示。小型电源输出 DC 48 V,功率模块功率等级从 800 W 到 3 000 W 有多种规格可供选择,可并联扩容。电池采用 48 V 电池组,保证 0.5~1 h 备电时长,可并联扩容以延长备电时间。考虑壁挂、抱杆时的可安装性,宜采用体积小、重量轻的锂电池组。本地 DC 48 V 供电解决方案为相对比较通用经济的供电方案。

3.3 本地 AC 与 DC 48 V 混合供电解决方案

部分 Small Cell 站点,存在既需要 AC 供电,也需要 DC 48 V 供电的情况。这种情况的产生一般是前期部署了部分通信设备,后期新增通信设备与之前的设备供电需求不同。例如:某站点前期部署了市电直接供电的 4G 设备(无备电,直接接入市电),后来站点扩容新增 DC 48 V 供电的 5G 设备,且两次部署的通信设备均需要备电。此种场景下采用的供电方案如图 3。图 3 a)为采用小型 DC



▲ 图1 本地AC供电解决方案(有备电)



▲ 图2 本地DC 48 V供电解决方案(有备电或无备电)

电源加独立逆变器方案,图3 b)则采用一个双路DC输出的电源系统,同时提供DC 48 V和DC 240 V输出。采用独立逆变器方案相对通用,但设备多、成本高。采用双路DC输出的电源系统方案不通用,但设备少、占用空间小。本地AC与DC 48 V混合供电方案的应用场景相对较少。

3.4 直流远端供电解决方案

部分Small Cell站点,难以在站点直接获取市电,例如:隧道或无市电覆盖的山顶等。此种场景下,可考虑采用直流远端供电方案,如图4所示。在局站端,增加一个DC/DC电源,将原有的48 V电源升至某一更高电压,例如280 V DC,然后通过电缆将电源拉到远端站点^[9]。对于可采用AC 220 V(兼容高压直流输入)的Small Cell通信设备,直接采用高压直流供电。对于DC 48 V供电的Small Cell通信设备,则需增加一个小型DC电源(此处小型DC电源与采用本地DC 48 V供电方案中的小型DC电源相同),转换为48 V后给设备供电。对直流远供电压没有严格规定,但业界比较通用的电压等级为DC

280 V,也可选用其它电压,如采用DC 270 V(标称240 V)、DC 380 V(标称336 V),与通信设备用的高压直流电源系统电压一致,此时远端AC供电的Small Cell通信设备需兼容这些电压输入。

直流远端供电方案,考虑拉远供电线缆成本及电能损耗的影响,对通信设备功耗、拉远距离均有要求,功

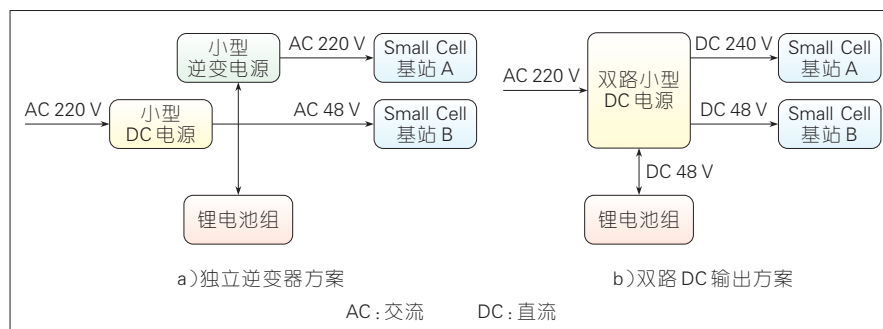
耗不易过大,一般不超过500 W,拉远距离不宜超过5 km。

3.5 以太网供电(POE)供电解决方案

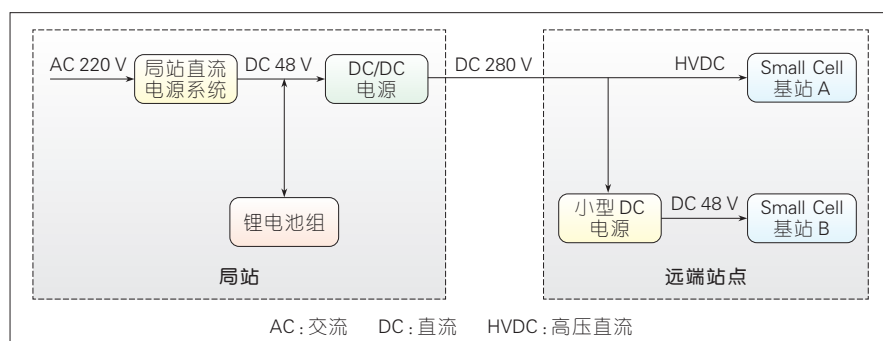
对于功率15 W左右的末端Femto Cell站点,可采用POE。POE供电是利用标准以太网传输电缆,同时传送数据和电功率的方案,如图5所示。在Small Cell站点上级的通信设备,通过以太网线将数据和DC 48 V电源混合传送到Small Cell站点。按照IEEE 802.3at标准,采用CAT-5e及以上网线时,POE最大供电功率可达25.5 W,供电电压范围42.5 V~57 V^[14]。

3.6 供电解决方案比较

不同供电方案适用于不同的应用场景,在建设成本、运维成本、供电可靠性等方面均有差异,需根据实际



▲ 图3 本地AC与DC 48 V混合供电解决方案(有备电)



▲ 图4 直流远端供电解决方案

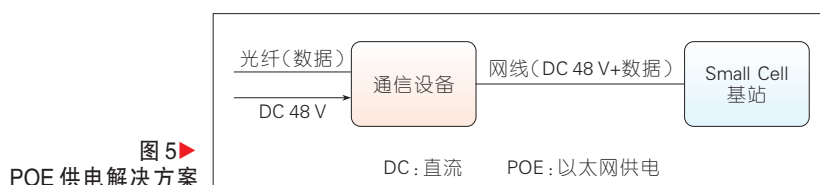


图5 POE供电解决方案

场景选取合适的供电解决方案。上述几种供电解决方案的比较如表2。相比而言,本地 DC 48 V 供电解决方案和 POE 供电解决方案是比较通用、经济的解决方案,推荐优先使用。

4 Small Cell 基站供电设备

4.1 设备形态

Small Cell 基站供电设备的形态,需要适应 Small Cell 基站复杂场景,满足前述零占地、零维护、易部署等需求。中兴通讯两种典型的 Small Cell 站点供电设备如图6所示。

图6 a)为 AC 220 V 输入,DC 48 V 1 500 W 或 3 000 W 输出的电源,图6 b)为 48 V 25 AH 锂电池模块。电源和电池均为自然冷却,IP66 防护等级,可应用在室外场景,支持多模块并联扩容。

4.2 安装方式

Small Cell 基站电源可适应挂墙、抱杆以及多模块旗装,如图7所示。可以满足零占地等需求,适应不同的应用场景。

5 Small Cell 基站供电方案的发展趋势

5.1 小型化、模块化

随着网络速率的提高,Small Cell 基站越来越靠近用户端。Small Cell 基站的体积越来越小,便于站址的选择与隐蔽,与之配套的供电设备也要更加小型化以便于安装。

模块化的供电设备可以实现任意并联,实现 Small Cell 基站分期部署及平滑扩容。模块化的供电设备可实现在 3G 基础上逐步叠加 4G,以及未来叠加 5G 业务的需求,实现通信基站的平滑演进及业务升级。

5.2 多样化

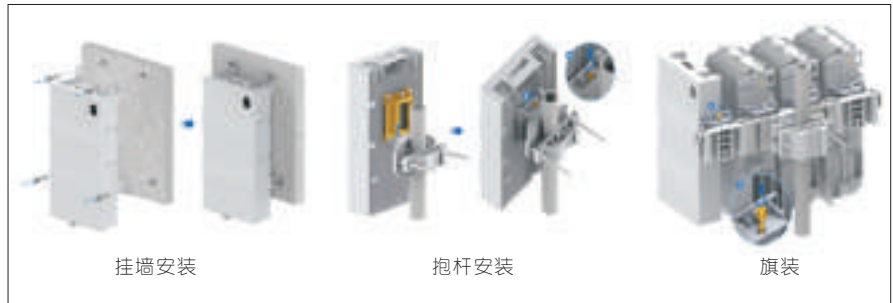
Small Cell 基站为实现定向覆盖、增强覆盖、盲区覆盖、边缘覆盖等多

▼表2 Small Cell 基站供电方案比较

方案类型	建设成本	运维成本	供电可靠性	说明	推荐程度
本地 AC 供电解决方案	低	中	中	适用 AC 供电设备	★★★
本地 DC 48 V 供电解决方案	低	中	高	适用 DC 供电设备	★★★★★
本地 AC 与 DC 48 V 混合供电解决方案	中	高	中	AC 与 DC 48 V 供电共存设备	★★
直流远端供电解决方案	高	低	高	本地取电困难	★
POE 供电解决方案	最低	低	高	室内,功率不超过 25 W	★★★★★

AC: 交流 DC: 直流 POE: 以太网供电

图6
Small Cell 基站供电电源系统



▲图7 Small Cell 基站电源的安装方式

种场景,以及复杂的现场部署要求,形成了多样化的基站形态。同时,供电功率从 10 W 到千瓦级,各有不同,形成了多样化的供电方案。未来 5G 的设备形态也会与现有的 3G、4G 设备有较大的不同,Small Cell 基站供电设备的形态将会随之变化。

5.3 高效节能

除了要求供电设备具有更高的转换效率(96%或更高),还需要供电设备能在基站负荷较低时下电,当负荷变高时重新启用,就像电子感应灯一样,以降低基站的电能消耗。当 Small Cell 基站下电后,由宏站或周边的 Small Cell 基站实现信号覆盖;当

相邻的宏站负荷变高时,Small Cell 基站就会被启用。

5.4 绿色

随着节能减排要求的提高,以及太阳能发电的成本越来越低,对于室外部署的 Small Cell 站点,将大量引入小型太阳能供电系统为站点供电,实现零排放。当前,对于部分市电获取困难的通信站点,太阳能供电已经具有较强的经济性。

5.5 共享

Small Cell 站点供电除了可实现 3G、4G 以及未来的 5G 基站共享供电外,也可以为不同的运营商的 Small

Cell 基站通信设备供电。随着物联网的发展, Small Cell 站点通信设备可升级为物联网基站, 同时也可物联网的终端共享供电, 如街边站的车联网传感设备、安防监测控制设备等。

5.6 智能、互联

软件定义电源(SDP)得到越来越多的应用, Small Cell 供电电源也将逐步实现软件定义, 实现多种电源制式的输出。同时, Small Cell 基站电源系统的能量接口和信息接口也将逐步符合未来能源互联网的标准, 可以接入能源互联网, 实现不同站点间的能量互联^[5], 进一步提高能效, 降低供电成本。

5.7 空间能量供应

物联网的发展推动了小型无线供电、空间微能量收集等技术的发展, 未来的小功率 Small Cell 站点, 供电方式可能会来自太阳能、无线供电, 甚至是空间微能量的收集, 摆脱

对电网供电的依赖, 大幅度降低取电成本。

6 结束语

移动通信的发展, 推动了 Small Cell 基站的大量部署。Small Cell 供电方案一方面适应 Small Cell 通信基站各种复杂场景的供电需求, 另一方面也符合电源的技术发展趋势, 为信息通信的发展提供了高效、绿色、灵活、可靠的供电解决方案。

参考文献

- [1] 朱龙明, 朱清华, 姚强. M-ICT 时代下的 5G 技术 & 创新[J]. 中兴通讯技术, 2016, 22(3): 52-56. DOI: 10.3969/j.issn.1009-6868.2016.03.012
- [2] 中兴通讯. Small Cell 基站系列产品[EB/OL]. (2017-05-10) [2017-06-02]. <http://www.zte.com.cn/china/products/wireless/Base-Station-Series/Small-Cell>
- [3] 中国通信标准化协会. 通信设备用直流远供电系统: YD/T 1817-2008[S]. 北京: 中国通信标准化协会, 2008: 7
- [4] IEEE SA. Local and metropolitan area networks—Specific requirements—Part 3: CSMA/CD Access Method and Physical

Layer Specifications Amendment 3: Data Terminal Equipment (DTE) Power via the Media Dependent Interface (MDI)
Enhancements: IEEE 802.3at-2009[S]. USA: IEEE, 2009: 10

- [5] 胡先红. 信息设备供电系统发展趋势[J]. 中兴通讯技术, 2016, 22(1): 42-45. DOI: 10.3969/j.issn.1009-6868.2016.01.011

作者简介



胡先红, 中兴通讯股份有限公司能源产品总工程师、能源产品规划首席专家, 广东省电源行业协会副会长, IEEE 会员; 对通信电源的架构和发展趋势有较深入研究, 主要负责中兴通讯能源产品的规划与架构方案设计工作; 已发表论文 10 余篇, 拥有中国专利 4 件。



刘明明, 中兴通讯股份有限公司能源研究院研发总工、研发部长, IEEE 会员; 现主要从事通信电源、混合能源技术研究及产品研发; 已发表论文 17 篇, 其中 10 篇被 IEEE 收录, 2 篇被 SCI 收录, 拥有专利授权 18 件。

基于运行大数据学习的复杂装备故障诊断技术及其典型应用

Fault Diagnosis Technology for Complex Equipment Based on the Learning of Big Operation Data and Its Typical Application

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0056-004

摘要: 认为通过从复杂装备运行特征大数据中挖掘出故障信息, 实现运行故障的智能诊断, 对保证复杂装备的安全和稳定运行具有重要意义。结合复杂装备运行大数据的特点与机器学习理论, 提出了基于运行大数据学习的复杂装备故障预测诊断方法, 实现了复杂装备运行特征大数据与运行故障的分层关联, 基于大数据分析的复杂装备运行故障特征提取以及基于模糊反向传播(BP)神经网络的复杂装备运行故障诊断。此外, 还将此技术应用到高速电梯的运行监测中, 开发了高速电梯急停故障大数据分析诊断系统, 很好地验证了该方法的有效性。

关键词: 大数据; 机器学习; 故障诊断; 复杂装备

Abstract: In this paper, we consider that it is significant to dig out the fault information from the big data of the complex equipment operation feature, and diagnose the operation fault intelligently for ensuring the safe and stable operations of complex equipment. By combining the characteristics of complex equipment's big operation data and the machine learning theory, a fault prediction and diagnosis method for complex equipment based on the learning of big operation data is proposed. In this way, some items are realized, including 1) the hierarchical correlation between the complex equipment's big operation data and the operation trouble; 2) the operation trouble feature extraction based on big data analysis; 3) the operation fault diagnosis based on fuzzy back propagation (BP) neural network for complex equipment. Moreover, by applying the proposed method to the high speed elevator's operation monitoring, the big data analysis and diagnosis system for emergency stop fault of high speed elevator is developed, and the effectiveness of the method is well verified.

Key words: big data; machine learning; fault diagnosis; complex equipment

刘达新/ LIU Daxin
裘乐森/ QIU Lemiao
王志平/ WANG Zhiping

(浙江大学, 浙江 杭州 310027)
(Zhejiang University, Hangzhou 310027, China)

键智能基础共性技术之一。

随着装备复杂程度的增加, 对于装备运行的监测往往存在装备监测点多、监测点的采样频率高、数据收集时间长等特点, 使得复杂装备故障诊断系统需要处理的运行数据量呈现爆炸性增长, 数百太比特级甚至拍比特级规模的大数据已经屡见不鲜。海量运行数据的产生, 意味着复杂装备故障预测诊断技术迎来了它的大数据时代^[1], 也对故障诊断技术的发展提出了新的挑战。由于大数据往往蕴含着很多在小数据量时不具备的深度知识和价值, 只有通过大数据的智能化分析和挖掘, 才能将其价值显露出来, 因此将大数据分析机器学习技术应用于装备运行过程的故障预测诊断, 通过从复杂装备运行特征大数据中挖掘出故障信息, 实现运行故障的快速诊断, 是近年大数据在装备领域的重要应用之一。

国际上在利用大数据技术进行产品性能监测、故障诊断和提升产品性能方面已经有了较为成功的应用

故障诊断技术是保证复杂装备安全、稳定运行的重要技术之一。故障诊断技术通过对装备运行状态的监测及其相应数据的分析处

理, 实现对装备运行故障的预测和诊断, 判断装备的状态是否处于异常状态, 或故障状态、劣化状态发生的部位或零部件, 预测状态劣化的发展趋势等, 已广泛应用于大型空分装备、汽轮机组、航空发动机、高速电梯等复杂装备的运行监测控制, 被列为中国智能装备产业重点发展的九大关

收稿时间: 2016-07-22

网络出版时间: 2017-07-05

基金项目: 国家自然科学基金资助项目(51475417, L1524002)

案例。例如：美国 GE 公司应用 2 万台飞机引擎上的各种传感器收集飞行数据，精确检测飞机的运行状况，甚至预测故障，并及时进行预防性维保，其对于某些型号的引擎能够提前 1 个月预测维护需求，预测准确率达到 70%；英国 Rolls-Royce 公司通过对服役于新加坡航空的 137 台飞机发动机使用大数据分析技术，成功降低了燃油消耗；美国特斯拉汽车公司利用大数据技术，采集和分析用户在驾驶时产生的加速度、刹车、转弯、电池充电和位置信息，以及用户驾驶习惯等大量数据，结合市场数据和企业生产、管理数据，进行产品及生产协同创新。

1 大数据与机器学习

狭义的大数据是指不能装载进计算机内存的数据。因此，对于每台电脑来说，当数据量增长到不能装载进内存的时候，此时形成的数据集合便成了大数据。学术界对于大数据进行了更为科学的定义^[1]：大数据是指无法在可接受的时间内用传统软硬件技术和工具对其进行感知、捕捉、管理、处理和服务的数据集合。针对大数据的特点，专家学者们相继提出了 3V、4V 甚至 5V 模型来进行描述，包括：Volume（体积大）、Velocity（速度高）、Variety（多样性）、Value（价值）、Veracity（真实性）等。

大数据分析主要分为简单分析和智能化复杂分析两大类^[2]。简单分析主要采用类似于传统数据库联机分析处理（OLAP）的技术和方法，用结构化查询语言（SQL）完成各种常规的查询统计分析；而大数据的深度知识和价值仅通过简单分析是难以发现的，通常需要使用基于机器学习和数据挖掘的智能化复杂分析技术才能实现。机器学习作为人工智能的核心，是获取大数据所隐含深度知识的关键技术，而且处理的数据规模越大，机器学习模型的效果一般来说会越好^[4]。

但是，由于大数据的海量、多维、多样、变化快等特性，使得小数据环境下基于内存处理的传统机器学习算法基本已不再适用于大数据下的应用问题。如何适应大数据处理的需求，这对机器学习的研究提出了新的挑战。大数据环境下的机器学习算法研究主要是从以下 6 个方面开展的^[5]：

（1）大数据分治策略与抽样。大数据分治与并行处理策略是大数据处理的基本策略，如何学习大数据的分布知识，将其用于优化负载均衡是一个亟待解决的问题。同时，需要依据一定的性能标准对样本空间进行筛选，剔除冗余和噪音数据，在不降低甚至提高某方面性能的基础上，最大限度地降低计算时间和空间的消耗。

（2）大数据特征选择。大数据包含的属性数和记录数巨大，导致处理算法的执行效率低下。如何采用降维和特征选择技术以降低大数据处理难度，是大数据特征选择技术迫切需要解决的问题。

（3）大数据分类。针对不同分类算法研究并行或改进策略成为大数据环境下分类学习算法研究的主要方向，诸如支持向量机分类、决策树分类、神经网络分类等方法。近些年，在人工神经网络基础上发展起来的深度学习技术^[6]，逐渐展现出其在处理大数据方面的独特优势。由于增加了隐层单元，多层神经网络比感知机具有更灵活且更丰富的表达力，可以用于建立更复杂的数学模型，能够深刻揭示海量数据里所隐含的复杂而丰富的信息，从而做出更为精准的预测。

（4）大数据聚类。聚类学习是最早被用于模式识别及数据挖掘任务的方法之一，并被用来研究各种应用中的大数据库。但是经典聚类算法在大数据环境下面临着数据量大、数据体积大、数据维度高等诸多挑战，改进现有聚类算法，提出新的聚类算法是大数据聚类研究急需解决的关

键问题。

（5）大数据关联分析。大数据的关联分析主要有并行和增量两种途径。并行关联分析算法通过将产生候选项集的过程并行化来提高运行效率，具有良好的加速比和伸缩性；增量方面则主要体现在序列模式挖掘上。

（6）大数据并行算法。将传统机器学习算法运用到大数据环境中的一个典型策略，便是对现有的学习算法并行化。目前大数据并行算法的研究已在一定范围内取得了一些进展，能实现对一定量级大数据的分析处理，如何建立更为高效的并行策略，实现大数据的高效处理仍将是当今的研究热点。

大数据环境下的机器学习，既不是单纯的机器学习，也不是单纯的大数据处理技术所能解决的问题，而是一个同时涉及机器学习和大数据处理的交叉性研究课题。

2 基于大数据学习的故障诊断方法

复杂装备运行过程中产生的海量特征数据蕴含了大量的故障信息，如何从复杂装备运行特征大数据中挖掘出故障信息，实现运行故障的快速诊断，对提高复杂装备的安全性，实现稳定运行具有重要意义。基于大数据分析的故障诊断可以在收集到复杂装备运行特征数据的基础上，应用聚类、决策树等机器学习算法对大数据进行知识挖掘，获得与故障有关的诊断规则，从而实现对复杂装备的故障预测和诊断。

目前，复杂装备故障诊断方法主要分为 3 类：基于数学模型的故障诊断、基于数字信号处理的故障诊断、基于知识的故障诊断^[7]。基于数学模型的故障诊断方法简单直观且易于理解，但需要深入分析复杂装备的结构、运行原理，对内部结构和运行原理过于复杂的装备难以建立数学模型；基于信号处理的诊断方法不需要

构建数学模型,容易实现,但只在复杂装备有明显的外部特征时才有效,不适用于那些没有明显外部特征的故障;基于知识的故障诊断具有良好的诊断效果和广泛的适用性,对复杂装备的智能化要求较高,是目前实现复杂装备智能故障诊断与健康维护的主要研究方向。

结合复杂装备运行大数据的特点与机器学习理论,我们提出了基于运行大数据学习的复杂装备故障预测诊断方法,包括复杂装备运行特征大数据与运行故障的分层关联,基于大数据分析的复杂装备运行故障特征提取以及基于模糊反向传播(BP)神经网络的复杂装备运行故障诊断3个方面。

(1)复杂装备运行特征大数据与运行故障的分层关联。

通过详细分析复杂装备运行故障出现的原因,建立运行故障原因的层次结构模型,分析和采集复杂装备运行特征大数据,建立复杂装备运行特征参数与运行故障原因的关联映射关系。

(2)基于大数据分析的复杂装备运行故障特征提取。

针对复杂装备运行特征大数据有较多噪声的问题,提出基于小波模

极大值的复杂装备信号降噪方法,对特征大数据进行处理,提高复杂装备运行数据的信噪比。针对复杂装备运行特征大数据维数多、规模大等问题,提出基于粗糙集属性约简的运行故障特征提取方法,筛选运行特征大数据中的冗余属性,获取与复杂装备运行故障相关的特征参数。

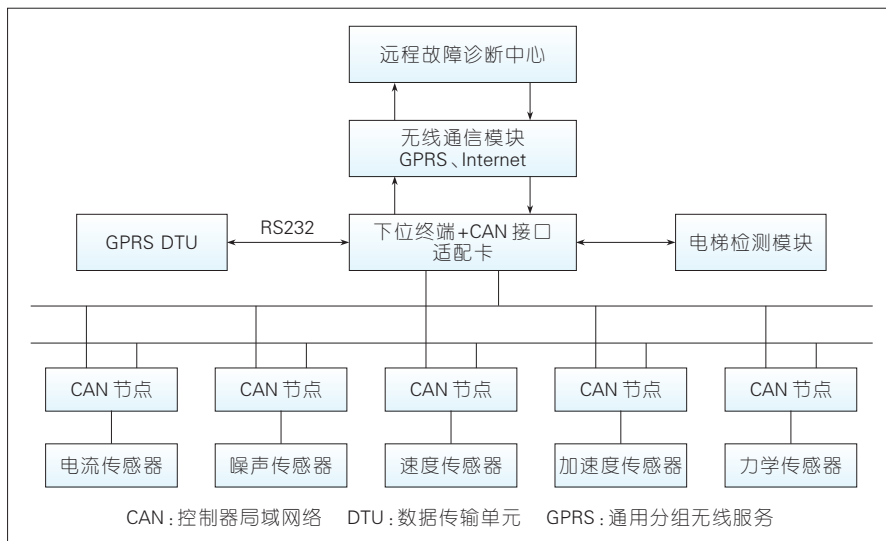
(3)基于模糊BP神经网络的复杂装备运行故障诊断。

基于复杂装备运行故障诊断的模糊BP神经网络构建,基于神经网络的复杂装备运行故障诊断方法和

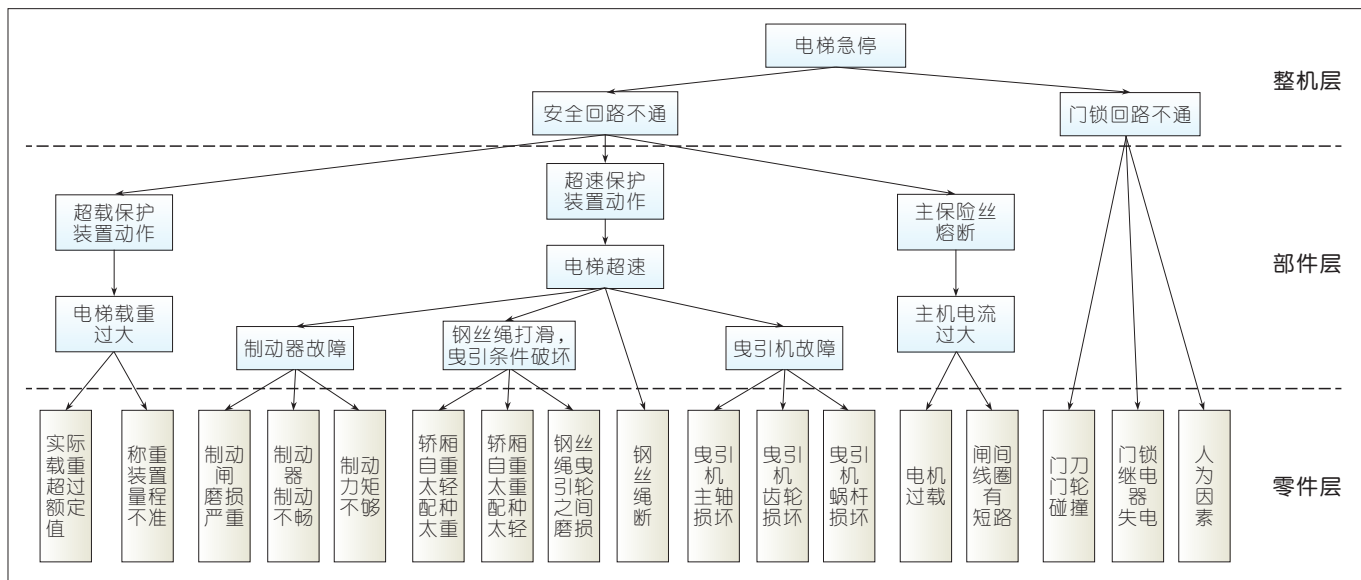
诊断过程的研究,建立复杂装备运行故障诊断模型,实现复杂装备运行特征大数据分析与基于知识的运行故障智能诊断。

3 典型应用案例

以高速电梯为研究对象,通过将研究成果应用到高速电梯急停故障的数据分析和智能诊断,结合高速电梯急停故障大数据分析与诊断系统的开发,验证所提方法的有效性。如图1,诊断系统通过收集高速电梯运行特征大数据,建立如图2所示的运



▲图1 高速电梯运行数据实时采集模块



▲图2 高速电梯急停故障的层次性

行特征参数与急停原因的分层关联映射。采用基于小波模极大值的高速电梯运行信号降噪算法去除运行特征大数据中的冗余信息,采用基于属性约简的高速电梯故障特征提取方法对运行大数据进行特征提取,获取高速电梯急停故障的特征参数,基于模糊BP神经网络构建了高速电梯急停故障诊断模型,如图3。

如图4所示,系统在江苏康力电梯企业额定速度7 m/s的KWG型高速电梯急停故障诊断中得到了应用验证,可在系统中查看KWG型高速电梯所有与急停故障相关的信息,有助于维修人员根据诊断结果对故障进行快速修复。

4 结束语

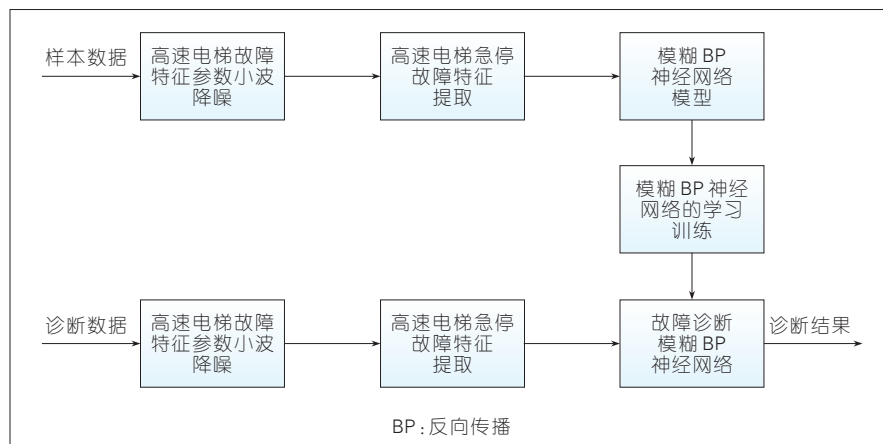
目前,中国在将大数据、人工智能等技术应用于复杂装备故障诊断

方面还处于理论研究阶段,在实际中主要以基于数学模型和基于信号处理的故障诊断方法为主。由于复杂装备的结构复杂,运行工况恶劣多变,难以建立精确的解析模型,运行信号也不能直观地反映装备运行故障的情况,因此传统的基于模型和信号处理的故障诊断方法已经不能适用于复杂装备的故障诊断。而随着物联网技术的发展,越来越多的装备制造商会通过传感器收集装备的运行数据,迅猛增长的装备运行特征数据对于分析其故障具有重要价值。因此,我们结合复杂装备运行大数据的特点,提出了基于运行大数据学习的复杂装备故障预测诊断方法,实现了复杂装备运行特征大数据与运行故障的分层关联,基于大数据分析的复杂装备运行故障特征提取以及基于模糊BP神经网络的复杂装备

运行故障诊断,应用于高速电梯故障预测与诊断,有效提高了复杂装备故障诊断的效率和准确性,取得了很好的应用效果。

参考文献

- [1] GRAHAM R D, GOLDSTON D, DOCTOROW C, et al. Big Data: Science in the Petabyte Era [J]. Nature, 2008, 455(7209):8-9. DOI: 10.1038/455001a
- [2] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012, 27(6):647-657
- [3] 黄宜华. 大数据机器学习系统研究进展[J]. 大数据, 2015, 1(1): 28-47
- [4] WANG Y, ZHAO X, SUN Z, et al. Peacock: Learning Long-Tail Topic Features for Industrial Applications[J]. Acm Transactions on Intelligent Systems & Technology, 2015, 6(4):1-23. DOI:10.1145/2700497
- [5] 何清, 李宁, 罗文娟, 等. 大数据下的机器学习算法综述[J]. 模式识别与人工智能, 2014, 27(4): 327-336. DOI:10.3969/j.issn.1003-6059.2014.04.007
- [6] HINTON G E, SALAKHUTDINOV R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786):504-507. DOI:10.1126/science.1127647
- [7] 王宏力, 侯青剑. 故障诊断方法现状与展望[J]. 传感器与微系统, 2008, 27(5):1-4. DOI: 10.3969/j.issn.1000-9787.2008.05.001



▲ 图3 基于模糊BP神经网络的高速电梯急停故障诊断



▲ 图4 高速电梯急停故障诊断结果

作者简介



刘达新, 浙江大学助理研究员; 主要研究方向为复杂产品数字样机与仿真、产品数字化设计与制造、计算机辅助检测等; 曾主持和参加国家自然科学基金项目3项; 已在重要学术期刊发表论文10余篇, 其中SCI/EI检索8篇。



裘乐淼, 浙江大学副教授; 主要研究方向为数字化设计制造、制造业信息化、数字化工厂与智能制造等; 曾主持国家自然科学基金2项、浙江省自然科学基金1项, 参与国家重大科技专项、“863”计划项目等10余项, 获国家科技进步二等奖1项、浙江省科技进步奖2项; 已发表SCI/EI论文30余篇。



王志平, 浙江大学硕士研究生; 主要研究方向为产品数字化设计制造、制造业信息化; 曾参与国家自然科学基金等项目2项。

BRAS SDN&NFV 的演进思路

Evolution of BRAS SDN&NFV

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0060-003

摘要: 认为宽带远程接入服务器(BRAS)与软件定义网络(SDN)、网络功能虚拟化(NFV)技术的结合是未来网络演进的方向,是实现未来弹性网络的关键。提出了作为电信网络边缘核心的 BRAS 设备在 SDN&NFV 技术大潮下的具体的演进思路: BRAS 与 SDN&NFV 技术的结合,采用控制与转发分离、软件与硬件解耦,实现弹性、可扩展、可编程的新型的宽带接入控制,为运营商提供灵活、高效、可靠的宽带接入服务。

关键词: BRAS; 宽带网络网关(BNG); SDN; NFV

Abstract: It is considered that the combination of broadband remote access server (BRAS) and software defined network (SDN) and network function virtualization (NFV) technology is the direction of the future network evolution, and is the key to realize the flexible network in the future. In this paper, we propose the evolution ideas of BRAS equipment in the SDN&NFV technology trend: combination of BRAS and SDN&NFV technology, separation of the control and forwarding, and decoupling of software and hardware. In this way, a new flexible, scalable and programmable broadband access control is realized, which can provide flexible, efficient and reliable broadband access to the operator.

Key words: BRAS; broadband network gateway(BNG); SDN; NFV

王怀滨/WANG Huaibin¹
王延松/WANG Yansong^{1, 2}

(1. 中兴通讯股份有限公司, 深圳 518057;

2. 移动网络和移动多媒体技术国家重点实验室, 广东 深圳 518057)

(1. ZTE Corporation, Shenzhen 518057, China;

2. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518057, China)

随着互联网业务不断丰富,特别是云计算、大数据、万物互联等业务的发展,对基础网络设施提出了更高要求,包括大带宽资源、高可靠性及灵活资源调度等。传统的电信网络在快速发展的互联网业务面前显得捉襟见肘。软件定义网络(SDN)和网络功能虚拟化(NFV)技术的涌现给网络演进注入了新元素与动力。宽带远程接入服务器(BRAS)设备位于城域网边缘,是用户实现各种业务的入口及策略执行点,也是整个互联网业务的核心节

点。BRAS 与 SDN、NFV 技术的结合是未来网络演进的方向,是实现未来弹性网络的关键^[1-2]。

1 现有 BRAS 部署及运营难题

现阶段, BRAS 设备是软硬件一体化的专用设备,是典型的控制与转发紧耦合网络设备。在设备规模应用及业务运营过程中,一系列问题逐渐浮现^[3]。

(1) 设备资源难以复用及共享。在单台设备环境中,经常出现转发面与控制面资源不协调的矛盾,部分高带宽要求节点,如承接光网城市高速接入用户区域,转发端口及带宽资源消耗殆尽,而用户会话及服务质

量(QoS)队列、访问控制列表(ACL)队列等控制面资源有余;部分节点承载着大量长时间在线业务,如 IP 多媒体子系统(IMS)、终端综合管理系统(ITMS)等,用户会话等控制面资源耗光,而空闲端口资源却十分丰富。此外,多台 BRAS 设备之间无法形成负荷分担,存在忙闲不均情况,导致投资资源未被充分利用。

(2) 设备能力提升与硬件强相关,升级困难。而在硬件升级方面,从 10 G 升级到 40 G, 40 G 再升级到 100 G, 100 G 升级到 400 G, 400 G 再到更高性能的硬件平台,每次提升都需大量的投资,且低能力设备及板卡可利旧率较低。在软件升级方面,业务功能实现与硬件相关,需要同时满足特定的硬件及软件升级才能达到相应的水平,给新业务部署带来极大的阻碍,拖延业务上线时间。

(3) 设备配置与设备底层操作系统(OS)强相关,基本采用命令行界面(CLI)方式,配置效率低,且不同设备需预置大量用户策略,运维工作量大。每台设备上线都需重新配置,无可重用性,对运营能力要求高。

收稿时间: 2016-03-20

网络出版时间: 2016-05-05

基金项目: 国家高技术研究发展(“863”)计划(2015AA016103)

(4) 设备实现融合业务难度大, 需硬软件同时升级。业务融合包括 BRAS 与业务路由器(SR)的融合, 与运营商级网络地址转换(CGN)的融合, 与防火墙(FW)的融合, 与深度报文检测(DPI)的融合等, 需配置相应的专用功能卡。此外, 软件功能需同时进行相应业务模块开发与整合, 增加软件版本复杂度, 且带来软件不稳定, 软件开发成本急剧上升, 软件升级周期长等问题。

2 BRAS 的转型之路

电信运营商互联网化转型需求迫切。业务云化成为了互联网业务发展的趋势, 这对底层网络提出了软件可编程要求。上层应用希望通过网络提供的可编程接口, 实现应用驱动的网络资源调度与管理, 促进增值应用的快速迭代开发。SDN 与 NFV 技术是实现网络转型的利器。SDN 强调转发与控制分离, 实现集中式管理, 提升网络运营效率。NFV 技术提倡采用通用 x86 服务器实现网络业务功能, 在降低硬件成本的同时提升软件可编程水平, 加快软件开发周期。

SDN、NFV 与边缘控制设备的融合(如图 1 所示), 是下一代智能边缘研发方向, 可解决当前遇到的难题与满足业务需求。虚拟宽带远程接入服务器(VBRAS)具备以下优势。

(1) 设备转发平面与控制平面分离, 打破当前设备封闭性, 实现网络功能与硬件的松耦合, 设备硬件及软件升级简单;

(2) 设备资源可虚拟化, 采用云技术进行承载, 可实现弹性伸缩, 并根据业务的具体需求, 动态地扩容或者缩减资源;

(3) 设备融合业务增强更加简单, 可采用软件模块实现专用的业务功能, 提升增值业务能力;

(4) 控制面集中管理, 可实现整网业务策略的统一部署, 提升业务上线速率, 增强市场竞争力;

(5) 理想情况下, 设备控制平面

与转发平面可形成相应的资源池, 实现弹性扩展及灵活调度, 提升资源利用率及高可靠性。

3 VBRAS 的部署

VBRAS 是运营商实现未来云网融合目标架构关键要素, 是业界研究热点。目前 VBRAS 主要还是在方案验证阶段, 预计 2017 年后才出现真正标准意义上的 VBRAS 产品。整体来看, VBRAS 有两条演进路线。

第 1 类 VBRAS 方案(如图 2 所示), 是传统的通信技术产业(CT)厂商普遍采用 SDN+NFV 的方案, 可称为“CT 线”。该方案优先实现转发控制分离, 控制平面采用通用 x86 服务器, 转发平面采用专用转发设备。该方案下的转发设备未来可基于白牌设备发展, 逐步演进至通用转发设备。CT 线暂不存在性能处理方面的问题, 但白牌设备的发展有待产业链进一步推动。该方案的典型特点是将传统的 BRAS 设备控制面与转发面

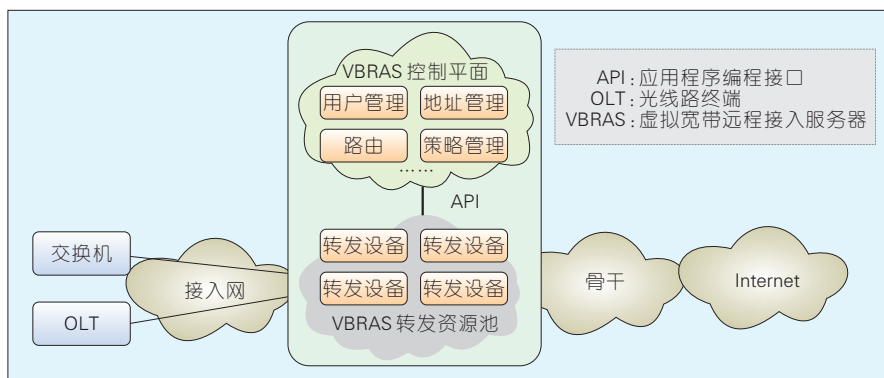
分离, 即控制面基于 NFV 架构实现虚拟化; 转发面基于专用的物理转发设备进行相关的业务转发。该方案具备以下优势。

(1) 转发面可以利用现有的 BRAS 设备, 充分利用现有资源, 从而保护运营商的现有投资。

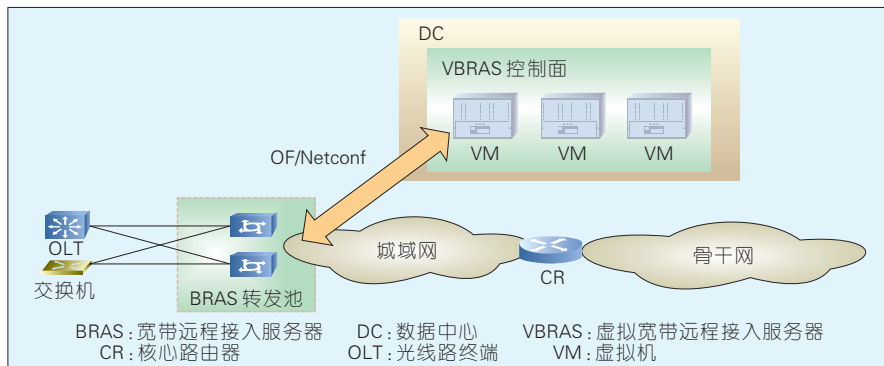
(2) 转发性能强。因为转发设备采用传统的物理 BRAS 设备转发, 转发性能强。现有的 BRAS 设备基本都是 100 G 及以上的平台, 支持每线卡 100 G 及以上的业务转发能力, 这是当前通用服务器达不到的。

(3) 多业务的叠加能力强。该方案实现控制与转发分离, 转发实体还是传统的 BRAS 设备, 整个网络架构并没有改变, 整体业务的处理流程没有变化。原来 BRAS 设备上的路由功能、运营商级网络地址转换(CGN)等多业务功能都很好的得到了继承。

(4) 转发时延小。用户的数据流量并没有进入数据中心, 在当前网络架构下, 转发路径更加扁平化, 用户



▲ 图 1 VBRAS 实现示意



▲ 图 2 VBRAS 分布式部署示意

业务流量转发路径短,转发时延小。

然而,该方案也有如下局限性。

(1)硬件平台不通用,需要专用的转发设备做转发,系统封闭。

(2)转发面虚拟化程度不高。因为是专用的硬件平台,很难纳入云平台的统一控制,从而很难在NFV统一架构下实现虚拟化。

第2类VBRAS方案(如图3所示),是信息技术产业(IT)厂商更喜欢采用SDN+NFV的方案,可称为“IT线”。该方案当前主要将BRAS的业务的整体功能运行在服务器上实现BRAS设备网元级的虚拟化,未来可基于软件功能分集,每个功能作为单独的虚拟机,从而实现BRAS业务组件级虚拟化。网元级虚拟化相比部署和扩容更加灵活。该方案的典型特点是BRAS的控制与转发都基于NFV架构实现虚拟化,部署在数据中心。方案优点是虚拟化程度高,具有弹性伸缩能力,部署更加灵活。

然而,该方案也有如下局限性。

(1)服务器转发性能不足,小字节报文处理性能较差,多业务处理能力欠缺等。尤其是BRAS业务最关注的用户多业务处理的QoS功能,目前服务器支持的性能和传统BRAS设备相差甚远,不能满足大用户量需求。

(2)多业务叠加能力弱。现在的BRAS设备实际是一个多业务融合的网关设备。BRAS设备上除了支持用户接入功能外,还有路由功能、组播功能、CGN功能、防火墙功能等。本

方案只是将其中的BRAS接入功能虚拟化,其他功能还没有虚拟化,从而导致多业务叠加能力差。

(3)组播支持能力弱。组播只能支持以太网上互联网协议(IPoE)用户的组播,不支持以太网上的点到点协议(PPPoE)用户的组播。

(4)网络时延长。用户的数据流量需要到数据中心处理完之后还需要重新回到城域网处理,数据流量转发经过的跳数多,转发时延大。在当前各种互联网业务层出不穷的时代,各种新业务对时延的要求越来越高,将所有流量全部送到数据中心的做法显得并不适当了。

(5)运营商网络部署困难,需要对现有网络进行大的改造。传统SR/BRAS作为城域网边缘的网络架构要发生大的变化是一个长期过程。

4 BRAS SDN&NFV 的演进思考

随着SDN及NFV大规模试点,业界表现出一定的理性。网络演进需同时兼顾传统网络部署及未来新业务需求。网络边缘设备需要同时支持丰富业务控制及流量高速转发的节点,以满足未来业务链发展。

(1)推动BRAS设备控制与转发分离。控制面基于服务器实现虚拟化,转发面使用专用转发设备,部署在城域网边缘,实现SDN架构的BRAS转发。运营商可在现有网络架构之下分享SDN&NFV带来的好处。

(2)推动控制平面集中部署及池化,实现控制平面的冗余及灵活扩展。同时推动互联网内容提供商(IDC)网络下移,在城域网网络服务提供点(POP)构建微型互联网数据中心(IDC),实施云平台在城域的延伸,并构建云化VBRAS池。

(3)基于云化思维构建城域虚拟网络服务提供点(vPOP),弹性部署虚拟深度报文检测(vDPI)等增值模块。部署城域业务链平台,提供基于用户的增值业务,实现新型业务收入。与此同时,推动通用转发设备产业链发展,逐步实现转发设备池化。

5 结束语

传统BRAS与SDN&NFV技术的融合是大势所趋,但大规模部署还有很长的路要走。结合具体的业务与应用场景,选择合适方案,分阶段部署将是未来BRAS与SDN&NFV技术结合的可行之路和长期演进过程。

参考文献

- [1] Broadband Forum. WT-345 Broadband Network Gateway and Network Function Virtualization [R]. USA: Broadband Forum, 2016
- [2] QUINN P, NADEQU T. Problem Statement for Service Function Chaining[R/OL]. [2016-04-05]. <http://www.rfc-editor.org/info/rfc7498>. DOI 10.17487/RFC7498
- [3] HALPERN J, PIGNATARO C. Service Function Chaining (SFC) Architecture[R/OL]. [2016-04-08]. <http://www.rfc-editor.org/info/rfc7665>. DOI 10.17487/RFC7665

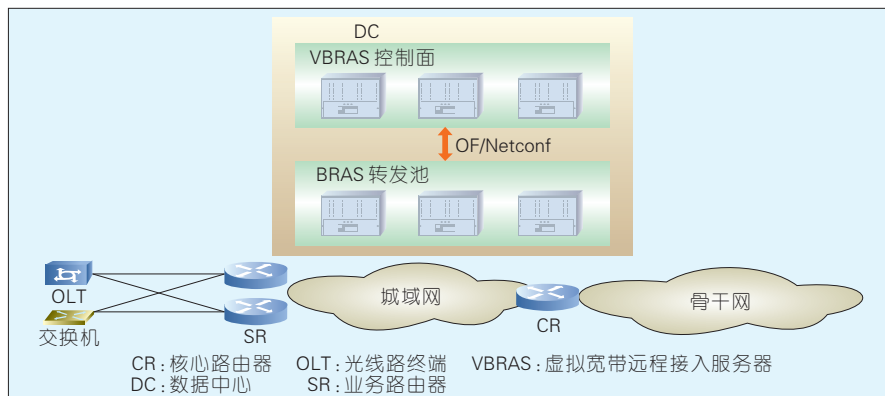
作者简介



王怀滨,中兴通讯股份有限公司承载网规划系统工程师;现从事有线BNG产品方面的研发规划工作;申请专利10项,发表文章3篇。



王延松,中兴通讯股份有限公司承载网预研总工,移动网络和移动多媒体技术国家重点实验室成员;现从事有线产品新技术预研工作,重点研究方向为SDN、NFV、IPv6、新型光网络等;先后承担国家“863”、CNGI等多个项目;已申请专利15项,发表文章5篇。



▲图3 VBRAS集中式部署示意

《中兴通讯技术》杂志(双月刊)投稿须知

一、杂志定位

《中兴通讯技术》杂志为通信技术类学术期刊。通过介绍、探讨通信热点技术,以展现通信技术最新发展动态,并促进产学研合作,发掘和培养优秀人才,为振兴民族通信产业做贡献。

二、稿件基本要求

1. 投稿约定

- (1)作者需登录《中兴通讯技术》投稿平台:tech.zte.com.cn/submission,并上传稿件。第一次投稿需完成新用户注册。
- (2)编辑部将按照审稿流程聘请专家审稿,并根据审稿意见,公平、公正地录用稿件。审稿过程需要1个月左右。

2. 内容和格式要求

- (1)稿件须具有创新性、学术性、规范性和可读性。
- (2)稿件需采用WORD文档格式。
- (3)稿件篇幅一般不超过6000字(包括文、图),内容包括:中、英文题名,作者姓名及汉语拼音,作者中、英文单位,中文摘要、关键词(3~8个),英文摘要、关键词,正文,参考文献,作者简介。
- (4)中文题名一般不超过20个汉字,中、英文题名含义应一致。
- (5)摘要尽量写成报道性摘要,包括研究的目的、方法、结果/结论,以150~200字为宜。摘要应具有独立性和自明性。中英文摘要应一致。
- (6)文稿中的量和单位应符合国家标准。外文字母的正斜体、大小写等须写清楚,上下角的字母、数据和符号的位置皆应明显区别。
- (7)图、表力求少而精(以8幅为上限),应随文出现,切忌与文字重复。图、表应保持自明性,图中缩略词和英文均要在图中加中文解释。表应采用三线表,表中缩略词和英文均要在表内加中文解释。
- (8)参考文献以20条左右为宜,不允许公开发表的资料不应列入。所有文献必须在正文中引用,文献序号按其在文中出现的先后次序编排。常用参考文献的书写格式为:
 - 期刊[序号]作者.题名[J].刊名,出版年,卷号(期号):引文页码.数字对象唯一标识符
 - 书籍[序号]作者.书名[M].出版地:出版者,出版年:引文页码.数字对象唯一标识符
 - 论文集中析出文献[序号]作者.题名[C]/论文集编者.论文集名(会议名).出版地:出版者,出版年(开会年):引文页码.数字对象唯一标识符
 - 学位论文[序号]作者.题名[D].学位授予单位所在城市名:学位授予单位,授予年份.数字对象唯一标识符
 - 专利[序号]专利所有者.专利题名:专利号[P].出版日期.数字对象唯一标识符
 - 国际、国家标准[序号]标准名称:标准编号[S].出版地:出版者,出版年.数字对象唯一标识符
- (9)作者超过3人时,可以感谢形式在文中提及。作者简介包括:姓名、工作单位、职务或职称、学历、毕业于何校、现从事的工作、专业特长、科研成果、已发表的论文数量等。
- (10)提供正面、免冠、彩色标准照片一张,最好采用JPG格式(文件大小超过100kB)。
- (11)应标注出研究课题的资助基金或资助项目名称及编号。
- (12)提供联系方式,如:通信地址、电话(含手机)、Email等。

3. 其他事项

- (1)请勿一稿两投。凡在2个月(自来稿之日算起)以内未接到录用通知者,可致电编辑部询问。
- (2)为了促进信息传播,加强学术交流,在论文发表后,本刊享有文章的转摘权(包括英文版、电子版、网络版)。作者获得的稿费包括转摘酬金。如作者不同意转摘,请在投稿时说明。

编辑部地址:安徽省合肥市金寨路329号国轩凯旋大厦1201室,邮政编码:230061

联系电话:0551-65533356,联系邮箱:magazine@zte.com.cn

本刊只接受在线投稿,欢迎访问本刊投稿平台:tech.zte.com.cn/submission

办刊宗旨

以人为本,荟萃通信技术领域精英;
迎接挑战,把握世界通信技术动态;
立即行动,求解通信发展疑难课题;
励精图治,促进民族信息产业崛起。

双月刊 1995 年创刊 总第 135 期
2017 年 8 月 第 23 卷 4 期

主管:安徽省科学技术厅
主办:安徽省科学技术情报研究所
中兴通讯股份有限公司
编辑:《中兴通讯技术》编辑部

总编:陈杰
常务副总编:黄新明
责任编辑:徐烨
编辑:卢丹,朱莉,赵陆
排版制作:余刚
发行:王萍萍
编务:王坤

《中兴通讯技术》编辑部
地址:合肥市金寨路 329 号凯旋大厦 12 楼
邮编:230061
网址:tech.zte.com.cn
投稿平台:tech.zte.com.cn/submission
电子信箱:magazine@zte.com.cn
电话:(0551)65533356
传真:(0551)65850139

出版、发行:中兴通讯技术杂志社
发行范围:全球发行
印刷:合肥添彩包装有限公司
出版日期:2017 年 8 月 10 日
中国标准连续出版物号: $\frac{\text{ISSN } 1009-6868}{\text{CN } 34-1228/\text{TN}}$
定价:每册 20.00 元,全年 120.00 元