



中兴通讯技术
ZTE Technology Journal
ISSN 1009-6868,CN 34-1228/TN

《中兴通讯技术》网络首发论文

题目：数据驱动的智能电信网络
作者：程强，刘姿杉
网络首发日期：2020-09-24
引用格式：程强，刘姿杉. 数据驱动的智能电信网络. 中兴通讯技术.
<https://kns.cnki.net/kcms/detail/34.1228.TN.20200924.1646.002.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

数据驱动的智能电信网络

Data Empowered Intelligent Communication Networks

(中国信息通信研究院, 中国 北京 100191)

(China Academy of Information and Communication Technology, Beijing 100191, China)

程强/CHENG Qiang

刘姿杉/LIU Zishan

摘要:

高质量的电信数据集是开展各类机器学习应用的前提。分析了电信数据所具备的特性及面向机器学习应用的要求,对数据驱动的电信网络智能化算法与应用进行了总结。电信网络拥有丰富的数据和计算资源,具有应用人工智能技术的巨大空间和潜力。利用人工智能技术赋能电信网络,有望实现网络智能部署、运维自治、智能管理和优化。

关键词:

数据驱动; 网络智能化; 人工智能; 机器学习

Abstract:

High-quality network data sets are the prerequisite for developing the machine learning applications. The characteristics of the network data and the requirements for applying in the machine learning application are analyzed in the paper, and then the data-driven intelligent network methods and use cases are summarized. The communication networks have a plentiful of data and computing resources, and thus have a huge potential for applying the artificial intelligence (AI) technology to develop the intelligent network. The application of the AI technology to empower the communication networks is expected to realize intelligent network deployment, autonomous operation and maintenance, intelligent management and optimization.

Key words:

data driven; intelligent network; artificial intelligence; machine learning

传统的网络运维与优化往往基于人工经验与建模的方法来进行,在当前复杂异构的网络中存在越来越多的局限性和低效性^[1],例如:网络参数众多,难以获得精确的网络模型;计算复杂度较高,很多优化问题属于 NP-hard 问题;缺乏对网络参数的实时获取与适应;难以获得多方统一和全局最优解。利用人工智能技术与机器学习算法,可以为网络提出基于数据驱动的实时感知、预测和管控能力,实现网络的自治运维、自主决策、智能优化、精准市场推广与服务创新等。

高质量的电信数据是实现网络智能化应用的前提,与传统大数据相比有很多特殊性^[2],对实时处理、统一管理、隐私保护和质量保证有很高要求。

1 面向智能化应用的电信网络数据

1.1 电信网络数据来源

电信网络中的数据可以从多个数据源来获得,根据电信数据的时空特点可以将其分为实体数据和时序数据;根据数据的域级可以将其分为网络级数据、链路级数据和设备/用户级数据。

(1) 网络级数据。

网络级实体数据包括站点基础信息(站点的归属、类型和配置、部署方式、布局信息等)、有源/无源的物理资源信息(物理设备类型、物理设备板卡信息等)、网元配置信息(小区

参数、无线站点工参配置信息、核心网网络配置数据等)、网络拓扑数据(实体、逻辑和业务拓扑数据等),以及供应商信息数据和网络知识库数据。网络级时序数据(网络状态数据例如网络资源利用率等)、网络性能数据(关键性能指标、服务质量、资源利用率和其他业务指标数据)、网络质量数据(丢包、抖动、时延和吞吐量数据等)、故障类数据(网络告警、事件和日志等)、运营类数据(信令、话务量等数据)。

网络级数据一般通过OTT或网络运营商的服务器进行收集。OTT服务器收集的数据包含大量的文字、用户个人资料、系统日志、流量内容和其他业务指标数据等。由运营商进行收集的典型的电信数据集有呼叫详细记录(CDR)数据、用户平面流量(UPT)数据和控制平面流量(CPT)数据等。CDR是电信网络中被研究最多的数据集,最初主要记录用户的语音和本文从而进行计费。CDR数据集主要包括用户标识、服务事件发生的时间、地点以及持续时间,还可能包括流量类型和大小等。CDR数据集结构性和可用性较高,但往往只提供连接中的终端信息,而且仅通过CDR数据集难以实现对用户行为的全面精准分析。UPT数据通常包含IP会话的开始、结束时间、设备/用户的标识符、服务类型、上下行流量以及位置信息等。CPT数据通常在网络的控制面进行收集,数据字段通常包括用户标识符、时间类型、基站ID和时间戳等。

(2) 链路级数据。

网络链路数据包括在无线侧通过导频信号或追踪技术进行收集的链路测量数据,例如信号参考接收功率(RSRP)、参考信号接收质量(RSRQ)、路径损失、信道状态信息、资源块(RB)承载量等、传输/接收状况、光纤传输的链路状态(如光信噪比、光纤中断事件信息等)。在无线网络中,无线测量报告(RMR)是常见的链路级数据,它是基于无线电测量所生成的用户终端(UE)侧报告,用来指导无线网络的接入操作和性能评测,数据字段通常包括用户ID、宽带信道质量指示、服务RSRP、RSRQ,有时也包含用户吞吐量、终端和所服务的网元设备的精确位置等信息。

(3) 用户/设备级数据。

用户/业务类实体数据包括用户基础信息(终端类型、终端配置信息、网络标识、用户设置、个人信息等)、应用类数据(社交偏好、健康状况等)。时序类数据包括系统log(软硬件故障和事件信息)、用户状态数据(移动轨迹、速度、温度等)、应用数据(应用流量、会话、使用情况等和应用质量数据等)。除了传统用户相关数据外,还包括来自智能网联汽车的车辆状态信息、道路感知数据和社交数据等,物联网中RFID数据、传感数据,以及来自智能电网、智慧家庭、智慧医疗、工业互联网等各类应用与设备的数据等。

1.2 电信网络数据特点

(1) 5V特性。

电信数据具备一般大数据的5V特性^[3]:大量(Volume)、高速(Velocity)、多元(Variety)、价值(Value)和真实(Veracity)。大量是指近年来随着5G、移动互联网和物联网等相关技术的发展,电信网络中的数据量正在呈现爆炸式增长,据预测,2021年中国的移动数据流量预计每个月达到9.3 EBs^[4];多元是指电信网络异构的网络结构和网络设备、多样化和个性化互联网应用,使得电信网络数据的来源、类型、结构等呈现多样化;价值是指来自亿万级网络设备和终端的电信数据呈现出低价值密度的特性,同时这些电信数据中包含很多冗余的、低精度和噪声数据;高速是指电信网络数据的生成和传输具有高速性,对电信数据的获取和处理需要迅速及时;真实是指对电信大数据的分析和处理可以反映真实世界的用户行为和网络状况等。

(2) 多维。

目前几乎所有的网络设备和移动终端都具备了记录位置信息和时间的能力,并嵌入越来越多的传感器,例如加速度计、指南针、温度计、陀螺仪和环境光传感器等,可以提供多维度的

数据。这些多维度的电信数据可以用来实现多种数据分析应用，包括描述类、预测类和指令类应用等。

（3）多边。

电信网络大数据可能来自于多个数据采集点，由OTT服务器提供的终端应用所产生的数据可能无法被运营商所采集，不同类型的电信网络数据因此被不同的数据拥有方所获取。

（4）多粒度。

由于不同传感技术精度的不同、应用信息粒度和采集时间颗粒不同等原因，使得电信大数据的属性呈现多种颗粒度。

（5）个性化。

电信数据中包含大量与用户个人信息相关的数据，例如身份、位置信息和轨迹等，这使得电信大数据分析和人工智能应用的过程中必须高度重视用户隐私问题。

1.3 电信网络智能化应用数据要求

电信数据在智能化应用中的一般流程包括数据获取、数据预处理（数据清洗、数据脱敏、数据聚合等）、数据存储、数据处理（数据标注、训练、推理等）和数据销毁等，在其应用的全生命周期内需要实现以下关键要求，来支撑高质量机器学习应用的开展。

（1）实时处理。

为保证智能电信网络的实时感知与推理对低时延传输的要求，随着5G网络的发展部署，网络超高密度的连接下，对来自海量设备的数据包进行实时、无损、均衡的采集、传输、存储、去冗余、压缩和标注等是需要解决的技术难点。

（2）规范统一。

不同电信网络智能化应用对于数据的需求在数据类型、实时性、范围和颗粒度等方面各不相同。随着电信网络架构区域扁平化，机器学习应用过程需要支持对数据的分布式处理^[5]，需要针对电信网络数据类型、特征、格式和颗粒度的多样性来进行统一的数据表示、数据融合、数据预处理、数据存储和数据应用。

（3）隐私保护。

电信网络数据包含了丰富的用户个人信息。为了保证数据安全和个人隐私，在电信网络智能化应用开展过程中一方面要确保电信数据智能化应用全生命周期内的个人信息保护，目前常用的技术包括数据扰动、差分隐私、同态加密、联邦学习等；另一方面要考虑隐私保护技术的开销以及对数据可用性和机器学习应用性能的影响。

（4）质量保证。

高质量数据集是保证机器学习在电信网络中应用性能的关键，现阶段电信行业仍缺乏公开高质量的标注数据集。与此同时，通信系统的数据往往具有一定的稀疏性，在遍历性上有所欠缺，在进行机器学习应用前需要对数据集进行质量检验等。电信数据集的质量保证需要包括准确性（标注正确率高于一定阈值，脏数据比率定于一定阈值等）、完整性（数据字段或属性缺失比率低于一定阈值、数据遍历网络）、一致性（网络特定上下文数据应符合一定的逻辑关系，不同网络层域的数据之间满足映射关系上的一致性）、时效性（数据的获取满足机器学习应用对于数据实时性的需求）等维度。

2 数据驱动的网络智能化算法与应用

根据应用的类型，智能电信网络应用包括感知类、预测类与决策类应用。

2.1 感知类应用

利用标注电信数据集来对网络操作行为模式和属性进行分类，对电信网络 and 用户的性能和行为进行回归、分类、归因和规则匹配等，可以用来解决电信网络中难以统一建模和以往依赖于人工经验的运维问题^[6]，适用的方法和应用主要包括：

（1）监督学习。

监督学习是指，由标注好的训练数据集中学到或建立一个模式（函数 / learning model），并依此模式推测新的实例（网络参数、用户特征等）。在电信网络中适用的模型和算法包括贝叶斯理论、决策树、随机森林、K-近邻，支持向量机、回归分析和（深度）神经网络等^[7]。在物理层，监督学习可以用来实现无线频谱管理、功率控制、调制格式识别，传输性能检测等，在网络层可以实现故障检测与分类、故障根因分析、智能路由、负载均衡等，在应用层可以实现流量分类、用户关系链发现、经营分析、精准营销和实时营销、个性化推荐等。

（2）序列模式挖掘。

序列模式挖掘是指，用于在包含时间或序列的电信数据集中发现频繁发生的有序时间或子序列模式。序列模式挖掘与关联规则挖掘之间的区别在于前者需要序列间的先后顺序，后者不关注事物之间的先后顺序。典型的算法包括Apriori算法、Agrawal和Srikant提出的广义序列模式概念（GPS）、等价类发现序列模式（SPADE）、PrefixSpan以及在此基础上进行改进的CloSpan算法等，可以用来进行网络资源利用率预测、用户行为分析、客服生命周期管理、Web访问预测等。

（3）聚类。

聚类是指，不需要标注，直接根据对象之间相似性来进行分组，典型的算法包括K均值、凝聚层次聚类和DBSCAN等，可以用来进行流量分类、用户内容推荐、移动性管理、异常检测和攻击防御等多类应用。

（4）知识图谱。

知识图谱用于描述电信网络中的各种概念、术语及其相互关系，可以用来进行电信网络的专家知识与运维经验的数字化。

2.2 预测类应用

预测类分析主要利用现有电信数据集来对未来的行为和趋势进行预测，是电信网络实现网络自适应调整与优化的基础，主要的算法和应用包括：

（1）时间序列分析算法。

时间序列分析算法，利用前期数值与后期数值的相关关系，建立包含前期数值和后期数值的回归方程，从而达到预测的目的，该算法包括自回归（AR）模型、移动平均（MA）模型、自回归移动平均（ARMA）模型和差分自回归移动平均（ARIMA）模型等。在电信网络中，可以用来实现网络长时流量预测、设备热迁移、异常检测、入侵检测等。

（2）回归算法。

回归算法用来确立目标特征之间关系函数，从而对连续值进行预测。电信网络中服务质量（QoS）参数、流量、网络资源利用率、无线路径损耗等数值往往是连续的，适合采用回归算法来进行预测分析。

（3）分类算法。

分类算法是指，将电信数据集中出现的列项分配到目标类别的过程。典型的算法包括朴素贝叶斯、贝叶斯网络、K最近邻（K-NN）、支持向量机（SVM）、决策树和随机森林等，适用的应用包括关键性能指标（KPI）预测、故障检测、传感器成簇、流量分类、负载均衡、智能缓存、智能路由等。

（4）随机建模算法。

电信网络应用较多的随机模型包括动态贝叶斯网络、马尔科夫模型、卡尔曼滤波器和扩展卡尔曼滤波器等，这些模型一般用于建模系统状态随时间转移的概率，例如马尔科夫模型中在给定当前信息的情况下，将来的状态则与过去的状态无关。通过收集一定数量和时间序列上的用户信息，来对随机模型进行参数估计，可以实现对未来用户和网络行为进行预测。

（5）深度学习算法。

随着网络计算能力的提升以及训练样本的增加，深度学习算法得到了越来越多的重视，文献

[8]对深度学习算法在无线网络中的应用进行了总结,常用的模型包括循环神经网络(RNN)、LSTM等,可以实现信号检验、异常检测、移动性预测、KPI预测等应用。

2.3 决策类应用

预测类分析利用电信数据集对网络状态进行分析,在特定目标下输出对网络最优的控制决策,从而实现网络的自适应调整与优化控制,主要的算法和应用包括:

(1) (深度)强化学习。

深度强化学习用来解决智能体在与环境交互的过程中通过学习策略以达到回报最大化或实现特定目标,环境中包含未知的变量或状态信息等。在电信网络中,可以在对网络参与无法完全已知的情况下实现基站智能节能、多信道动态接入、功率分配、智能缓存等应用。

(2) 博弈论。

在存在多用户的电信网络中,博弈论可以针对具有竞争或合作性质的个体之间研究它们的优化策略,在电信网络中已经出现了相关应用,例如多用户的频谱感知、多用户合作的内容分发等。

3 结束语

基于 AI 技术的电信网络智能化已经成为目前通信行业的研究热点,其核心为数据驱动的网络感知、预测、推断和决策能力的提升。在网络智能化的实现过程中,高质量的电信数据是进行机器学习应用的先决条件。通过对电信数据进行实时处理、规范统一、隐私保护和质量保证,有望在电信网络中实现智能运维、智能网络优化、智能网络控制优化、用户个性化服务与创新等内生智能能力。

参考文献

- [1] 尤肖虎, 张川, 谈晓思, 等. 基于 AI 的 5G 技术——研究方向与范例[J]. 中国科学, 信息科学, 2018, 48(12): 1589-1602. DOI: 10.1360/N112018-00174
- [2] CHENG X, FANG L, YANG L. Mobile big data based network intelligence [J]. IEEE Internet of Things Journal, 2018, 5(6): 4365-4379. DOI: 10.1109/JIOT.2018.2864093
- [3] CHENG X, FANG L, YANG L, et al. Mobile big Ddata: the fuel for data-driven wireless [J]. IEEE Internet of things Journal, 2017, 4(5): 1489-1516. DOI: 10.1109/JIOT.2017.2714189
- [4] CISCOVNI Complete Forecast Highlight预测报告
[EB/OL].(2020-00-00).https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-for-ecast-highlights/pdf/China_2021_Forecast_Highlights.pdf
- [5] PARK J, SAMARAKOON S, BENNIS M, et al. Wireless network intelligence at the edge [J]. Proceedings of the IEEE, 2019, 107(11): 2204-2239. DOI: 10.1109/JPROC.2019.2941458
- [6] WANG T, WANG S, ZHOU Z H. Machine learning for 5G and beyond: from model-based to data-driven mobile wireless networks [J]. China Communications, 2019, 16(1): 165-175
- [7] SUN Y, PENG M, ZHOU Y, et al. Application of machine learning in wireless networks: key techniques and open issues [J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3072-3108. DOI: 10.1109/COMST.2019.2924243
- [8] MAO Q, HU F, HAO Q. Deep learning for intelligent wireless networks: a comprehensive survey [J]. IEEE Communications Surveys & Tutorials, 2018, 20(4): 2595-2621. DOI: 10.1109/COMST.2018.2846401

作者简介

程强, 中国信息通信研究院技术与标准研究所主任工程师、ITU-T ML5G焦点组副主席、中国人工智能产业发展联盟电信项目组组长; 主要从事电信人工智能、宽带网络等技术和标准

研究，起草多项宽带和网络智能化方向行业标准和技術白皮书。

刘姿杉，中国信息通信研究院技术与标准研究所工程师；主要从事电信人工智能、先进计算等技术和标准化研究。

