

基于随机森林的区域性通信质量感知系统探索

黄健文, 丁 奕, 欧阳辉, 苏丽裕, 匡磊怀

(中移互联网有限公司, 广东 广州 510000)

摘要:移动通信技术更新换代快, 所以当前 2G/4G/5G 多代通信技术并存运行, 且随着移动互联网的快速发展, APP 业务场景纷繁复杂。在这种背景下, 当出现区域性通信质差时, 运营商使用常规预定义检测模型有较大的场景局限性, 无法快速定位问题与修复故障。为此, 需要建设一个高时效且可以适应多样化业务场景的感知系统。文章先介绍了用户上报质差后检测与预定义模型预测质差两种模式的局限性, 再探索了系统的分层架构, 再深入探索质差感知时效性的关键提升, 最后再探索了感知模型的优化提升。该文探索的系统方案, 实现了对通信质差区域的近实时发现, 支撑了运营商快速响应并执行维修闭环。

关键词:手机信令; 通信质量; 大数据; 机器学习; 随机森林

中图分类号: TN929.5

文献标识码: A

文章编号: 2096-9759(2023)06-0217-03

Research of Communication Quality Perception System Based on Machine Learning

HUANG JanWen, DING Yi, OU YANG Hui, SU LiYu, KUANG LeiHui

(China Mobile Internet Co., Ltd, Guangzhou, 510000, China)

Abstract: The mobile communication technology is updated rapidly. Currently, multiple generations of communication technologies such as 2G/4G/5G coexist. With the rapid development of the mobile internet, the APP business scenarios are complex. In this context, when the regional communication quality is poor, the use of conventional pre-defined detection models by operators has greater scenario limitations, and can not quickly locate problems and repair faults. It is necessary to build an awareness system with high efficiency and can adapt to diverse business scenarios. This paper first introduces the limitations of the two models of detecting and predicting the quality difference after the user reports the quality difference. Then explore the hierarchical architecture of the system. Further explore the key improvement of perceived timeliness of poor quality. Finally, it explores the optimization and improvement of perception model. The system scheme explored in this paper realizes near-real-time detection of poor communication quality areas, and supports operators to respond quickly and implement maintenance closed-loop.

Keywords: big data; mobile phone signaling; communication quality; machine learning; random forest

1 引言

常规的区域性通信质量感知方法依赖于人工检测。即需要区域内用户大规模报障, 然后由客服部门总结判断是该区域的共性问题, 再通知运维人员上门检测后才能确定。这种常规做法的局限在于假若报障投诉数量不够, 就不能认定为区域性共性质差, 而且时间上也严重滞后。

目前行业内也已经有了有一些通信质量检测的探索方案^[1-2]。这些方案是把大批量的手机信令, 批量导入 hive 这样的数据仓库, 再根据预设阈值与计算规则, 来批量计算通信质量分数。数仓通常按天分区, T+1 时延执行任务调度。这些方案确能检测到一些特定场景下的通信质差现象。但时效性还有很大的提升空间。另外当前 2G、4G、5G 多代并存, 且手机的使用场景非常多样化, 预定义的检测模型有较大的场景局限性。

针对上述问题, 依靠大数据实时计算与实时查询技术及机器学习技术, 本文设计了一套具有高时效性的智能感知区域通信质量的系统。结合 B 域与 M 域数据后, 能实现对小区、网络维度的汇聚分析, 及对特定人群的汇聚分析等功能。

2 系统技术分析

通信质差感知系统囊括了信令数据采集, 感知模型训练, 质差感知预测, 感知汇聚分析等能力。通过对实时采集、数据湖和实时计算技术的运用, 设计了一个手机产生信令数据即可被实时采集, 并实时使用感知模型预测取得感知结果的高

时效感知体系。系统的分层架构如图 1 所示:

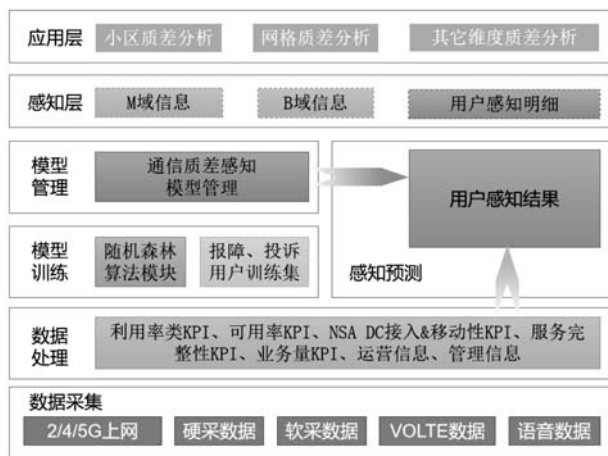


图 1 通信质差感知系统功能架构图

数据采集层: 基于 filebeat 实时采集信令数据, 包括 2/4/5G 上网、LTE 硬采数据、软采数据、VOLTE 数据以及语音数据。

数据处理层: 负责信令数据的清洗处理, 生成两百多个 KPI 指标^[1]。本层实现从实时数据汇聚到有考核意义的 KPI 数据, 如 UE CONTEXT 掉线率、TAU 时延、Attach 成功率、DNS 解析时延等两百多个指标。

收稿日期: 2023-03-12

作者简介: 黄健文(1978-), 男, 广东人, 教授级高工, 博士, 研究方向: 移动通信; 丁 奕(1989-)男, 广东人, 学士, 研究方向: 电子信息科学与技术; 苏丽裕(1988-), 女, 福建人, 高级工程师, 硕士, 研究方向: 计算机科学与技术; 欧阳辉(1987-), 男, 广东人, 硕士, 研究方向: 计算机科学; 匡磊怀(1995-), 男, 湖南人, 硕士, 研究方向: 深度学习。

表 1

时间维度	城市 ID	手机号码	小区 ID	业务 ID	指标 1	指标 N
20210312	86010	13800138002	10332	2143	0.68	102400
20210312	860759	13800138003	23332	3345	0.99	204800
.....

模型训练层:负责通信质差感知模型的训练。由于当前 2G、4G、5G 多代组网并存,且手机的使用场景非常多样化,本文选择了能对特征集灵活组合使用的随机森林算法。

模型管理层:负责质差感知模型的持久化与管理。训练出来的模型以文件的形式存储,供预测时加载使用。

感知层:负责对用户的最新指标记录做质差感知预测。预测模块根据记录中的地市维度与组网代次选取感知模型实例,执行质差感知预测,并将结果写入实时数仓 Clickhouse。

感知层:负责关联用户感知明细与 M 域 B 域信息。包括通信质差结论数据、地理维度数据以及用户行为数据等。

应用层:负责通信质差的多维分析^[4]。其中对小区与网格的质差分析可以快速发现区域的共性质差问题,从而实现快速通知运维人员上门检测并确定问题。

3 数据时效关键提升

本通信质差感知系统的处理流程如图 2 所示:

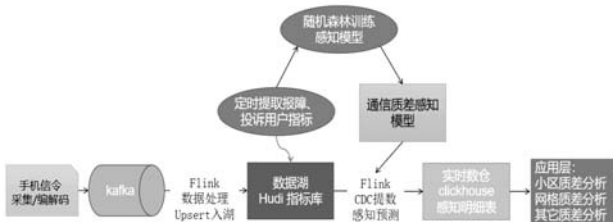


图 2 通信质差感知系统数据流程图

数据流在时效性的优化上做了以下关键提升:

使用 Flink 实时消费 kafka 的增量消息,统计 KPI 指标,无时延;后面使用 Flink 以 CDC 模式实时读取 hudi 指标进行感知预测,也无时延。

Upsert 模式入数据湖,可实现实时增量入湖,无时延。避免了像常规方案里累积批量数据再导入数据仓库的等待时延。

指标库新增数据模块使用了 CDC(改变数据捕获)规范设计,在对源表进行 INSERT、UPDATE 或 DELETE 等操作时变化的数据被保存在变化表中。同时为了避免侵入式 CDC 操作给指标库带来性能影响,采用了基于日志的 CDC 这种非侵入式模式。结构图如下:

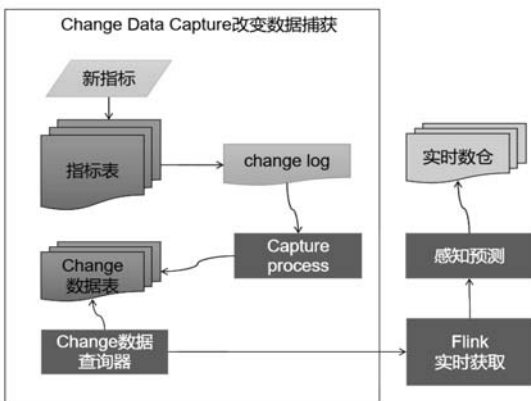


图 3 CDC 规范示意图

通信质差感知系统的常规处理流程时延评估模型如下:

$$T_{c时延} = t_{采集} + t_{批量时间窗口} + t_{信令入在} + t_{信令批处理} + t_{模型训练} + t_{批量感知预测} + t_{结果入在} + t_{批量感知构建} + t_{应用分析}$$

本系统的通信质差感知系统的常规处理流程时延评估模型如下:

$$T_{n时延} = t_{采集} + t_{kafka发布} + t_{信令处理} + t_{指标入库} + t_{模型训练} + t_{感知预测} + t_{结果入CK} + t_{应用分析}$$

经一系列提升,常规 T+1 模式质差感知的时延区间为[4, 28+]小时,实际值在区间内随批数据量有波动;经提升后成为了近实时模式,时延区间为[1, 60)秒。

4 模型提升

通信质量感知的核心是感知模块。常规做法是使用逻辑回归算法,经研究对比,选择了随机森林算法来提升模型。

按逻辑回归的算法,模型里训练出了一组权重 w^T ,作用于特征因子 KPI 指标数组,最后得到概率 P。但模型上的特征因子 KPI 指标固定有 213 个,实际上每次质差的场景,只对应了其中的几个,其余特征因子也固定参与预测不够合理。

$$P = \frac{1}{1 + e^{-(b + w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n)}} \quad (1)$$

随机森林的强泛化特性更切合用户组网特点与信令的多样性特点。它能随机组织特征因子子集组成决策树,模型中的决策树包含了各种特征因子,但又通过投票机制突出最优因子的作用从而消除了噪音。后经分析感知模块训练出来的模型文件,决策树上的特征因子汇总后涵盖了 2/4/5G 多代组网及全场景的 213 个指标。

在模型训练中,历史报障与投诉用户群标记为质差标签,客户调查中高满意度用户标记为质优标签^[5]。以地市和组网类别降维后再分别训练模型,并动态选择对应模型应用到不同的类别的用户的感知预测中去。

由于质差诱因指标集只占指标总集的一小部分,为了进一步优化让决策树聚焦到有价值的指标上,本系统在特征工程上引入了方差分析,去掉取值变化小的特征^[6]。指标值是离散型变量,如果在训练样本中某特征很大比例的记录的取值都是同一个值,那就可以认为这个特征作用较弱。通过计算低方差来筛选排除此类特征。

方差分析公式:

$$avg = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (2)$$

n 是数据个数, $x_1, x_2, x_3, \dots, x_n$ 表示这组数据具体数值。

$$s^2 = \frac{(avg - x_1)^2 + (avg - x_2)^2 + (avg - x_3)^2 + \dots + (avg - x_n)^2}{n} \quad (3)$$

经验证对比,使用逻辑回归算法的质差召回率为 49%,改用随机森林后召回率有 15 个点的增长,再经过一系列模型提升后,召回率再有 4 个点的提升。

表 2 模型平均准确率与召回率对比

模型	结果类型	质差准确率	质差召回率
逻辑回归	质差用户	51%	49%
随机森林	质差用户	65%	64%
随机森林模型提升后	质差用户	69%	68%

5 项目成效

本系统进行了一系列的数据时效性关键提升后,实现了秒级别的通信质差发现。随机森林模型的应用与提升也使召回率从 49%提升到了 68%。本系统上线后,覆盖全省 1.6 亿手机用户,日处理信令近 3000 亿条,平均每日能发出质差小区告警达 60 次,累计已超 2.16 万次,后经检测确认有超 1.46 万次小区

信息通信行业在车联网产业融合创新发展中的可为与作为

陈镠汀

(湖北省网络信息安全技术管控中心,湖北 武汉 430000)

摘要:当前 5G、大数据、云计算等新一代信息通信技术正深刻改变着各行各业。车联网产业是新一代通信网络、智能汽车、智慧交通融合发展的新兴产业。文章首先概括了目前我国车联网发展现状,进而介绍了车联网技术架构和关键技术(蜂窝车联网, C-V2X),探讨了 C-V2X 技术优势以及 LTE-V2X、NR-V2X 演进路径。最后,重点阐述了 5G 对车联网新型信息基础设施建设起到的重要推动作用,描述了我国“5G+车联网”发展路径;介绍了现阶段基础电信企业、电信设备运营商、互联网企业工作成果,指出了未来信息通信行业企业在车联网生态中的发展方向,并从发展和安全两个维度对目前产业政策进行了总结归纳。

关键词:车联网;融合发展;信息通信

中图分类号:F426.47

文献标识码:A

文章编号:2096-9759(2023)06-0219-03

The effects and features of the Information and Communication Industry in integrating development of the Internet of Vehicles industry

CHEN Liuting

(Network Information Security Technology Administration Center of Hubei Province,Wuhan 430000,China)

Abstract:The current new generation information and communication technologies such as 5G, big data, and cloud computing are profoundly changing various industries. As an emerging industry,The Internet of Vehicles industry integrates next-generation communication networks, intelligent vehicles, and smart transportation. Firstly, summarize the current development of the Internet of Vehicles.Then introduce the architecture and key technologies of the Internet of Vehicles (Cellular Internet of Vehicles, C-V2X), inquire into the advantages of C-V2X technology and the evolution paths of LTE-V2X and NR-V2X. Finally, focus on 5G which play an important role in the IoV critical information infrastructure , describe the development path of "5G+Internet of Vehicles" ; introduce the work achievements of telecom enterprises and ISP, point out the development direction, summarize the industrial policies on development and cyber security.

Key words: Internet of Vehicles industry;Integrate development;communication industry

1 引言

车联网是汽车、电子、计算机、通信等多技术领域深度融合的新型产业。发展车联网产业,有利于提升汽车网联化、智能化水平,对推动制造强国和网络强国建设,实现高质量发展具有重要意义。

2 我国车联网发展现状

2018 年 12 月,工业和信息化部印发《车联网(智能网联汽

车)产业发展行动计划》。从技术研发、标准制定、测试验证等方面刻画了我国车联网产业的生态图谱。2020 年 2 月,十一部委联合发布《智能汽车创新发展战略》,提出了 2035 年-2050 年中国标准网联汽车体系全面建成、智能网联汽车强国愿景逐步实现,充分满足人民日益增长的美好生活需要的三大远景目标。2021 年 3 月,国务院发布《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》,制定了“积极稳妥发展车联网”等任务目标,强化了车联网的战略地位。

收稿日期:2023-02-20

作者简介:陈镠汀(1991-),女,湖北武汉人,研究生,中级工程师,主要从事信息通信行业网络与信息安全管理相关工作。

确有软硬件故障。有效提升了质差问题的解决时效,降低了的用户报障与投诉数量。

6 结语

本通信质量感知系统得益于大数据实时计算框架、数据湖、实时数仓以及随机森林机器学习的应用,已具备了较好的应用效果。但也仍存在提升的空间,比如流程长而复杂,多个实时框架串联,对每一环节的稳定运行要求严苛。未来系统仍要在健壮性与高可用方面不断迭代优化,提升感知模型的精准度与准确性。

参考文献:

[1] 王瑞,曹颜辉,朱宁,梁星,张利峰.大数据分析在移动通信网

- 络优化中的应用[J].通信电源技术,2021,38(01):171-173.
- [2] 谢华,宋俊辉.大数据分析在移动通信网络优化中的应用研究[J].现代信息科技,2019,3(03):50-51+54.
- [3] 李会志,袁超伟.移动通信网络 KPI 指标分类方法研究[J].信息网络安全,2014(12):56-60.
- [4] 汪巍.移动通信网络质量分析与呈现系统的研究与实现[D].华中科技大学,2014.
- [5] 颜敏,高随祥,杨文国.移动网络用户体验质量评价的模糊层次分析法[J].网络新媒体技术,2017,6(03):1-7.
- [6] 周炎龙,孙广路.双重代价敏感随机森林算法[J].哈尔滨理工大学学报,2021,26(05):44-50.