

基于机器学习算法的LTE高投诉 小区预判方法

Prediction Method of LTE High Complaint Cell Based on Machine Learning

曹丽娟,程新洲,徐乐西,张 涛,贾玉玮,成 晨(中国联通网络技术研究院,北京 100048)

Cao Lijuan, Cheng Xinzhou, Xu Lexi, Zhang Tao, Jia Yuwei, Cheng Chen (China Unicom Network Technology Research Institute, Beijing 100048, China)

摘 要:

将用户投诉映射到网络性能问题,并利用机器学习算法建立4G KPI与用户投诉之间的关联,构造基于网络性能指标的用户投诉预警模型。以用户感知层面的大数据分析结论为抓手,提升网络优化和网络运维的质量,并提升其实时性和准确性。将输出的决策树模型以静态规则部署在系统中,可将预判的疑似问题小区名单直接送达建维优相关人员,实现了高投诉小区问题解决的闭环以及问题的可追踪可回溯。

关键词:

大数据;机器学习;LTE;用户投诉;KPI

doi:10.12045/j.issn.1007-3043.2020.05.004

文章编号:1007-3043(2020)05-0018-04

中图分类号:TN929.5

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

It sets a mapping from user complaints to network performance problems. The correlation between 4G KPI and user complaints is established by using machine learning algorithm, and an warning model of user complaints based on network performance indicators is constructed. Based on the conclusion of big data analysis of the user perception, the operator can improve the quality of network optimization and network operation and maintenance, and improve the realtime and accuracy. By deploying this strategy in the network with static rules, the suspected problem cell list of high complaint can be directly exported to the relevant personnel. In this way, the closed-loop of problem solving is realized and the problem is traceable.

Keywords:

Big data; Machine learning; LTE; User complaints; KPI

引用格式:曹丽娟,程新洲,徐乐西,等.基于机器学习算法的LTE高投诉小区预判方法[J].邮电设计技术,2020(5):18-21.

0 引言

移动通信网络技术的不断发展催生了日益复杂的移动网络环境,对运营商的网络优化提出了新的挑战。

用户投诉数据是用户意见的直接反馈,运营商通过多渠道获得了大量用户投诉数据,其中蕴含了大量

有价值的数据和信息^[1]。通过投诉处理可以及时有效地了解网络与业务中的不足,第一时间跟进处理,从客户投诉/用户感知入手,提升网络整体支撑保障能力与客户感知,提升网络优化效率^[2]。然而,当前运营商处理用户投诉主要停留在投诉事后的处理过程,缺乏投诉的预判和防范,投诉处理过程相对复杂,障碍申告和网络优化的实时性不强。针对传统的运营商投诉处理过程中的不足和局限性,本文提出了一种基于机器学习的系统化和自动化的4G网络投诉预判分析系统。

论文组织如下:首先介绍了运营商投诉处理的传统流程,归纳了这种方法可能存在的缺陷;随后论述

基金项目:国家重点研发计划(2018YFB1800800);工业和信息化部大数据产业发展试点示范项目(融合异构数据及深度学习的民生大数据创新应用试点示范)

收稿日期:2020-04-08

了本方案的核心流程,包括数据采集及预处理、数据入库及中间表建立、决策树构建剪枝以及可视化的流程;最后从具体实施的角度,介绍了此方案的部署和应用,并对本方案进行了陈述及总结。

1 传统的运营商投诉处理流程

投诉处理是局部网络优化的重要手段,是改善局部网络覆盖质量的重要参考,因此投诉处理在网络优化工作中有重要的指导意义。运营商的典型投诉处理流程可以简单描述为:

a) 投诉信息收集:一线客服平台收集用户的咨询及投诉。用户将投诉的位置、具体现象和表征进行描述,一线客服人员对这些现象和表征进行提取汇总,获取具体、详尽、完备的投诉信息。

b) 电子运维系统派发工单:客服人员将投诉详单录入电子运维系统,并对用户投诉的问题进行初步分类(如网络制式分类:2G/3G/4G;投诉类型分类:服务相关投诉/通信质量相关投诉)。筛选出通信质量相关投诉,并通过电子运维系统对已发生的用户投诉进行工单派发。

c) 投诉处理:相关网络的运维人员将用户投诉映射到网络问题,结合设备告警等进行投诉原因定位。结合投诉地点的实际情况,对可能存在的问题进行实地排查和测试,进而对相应的问题(网络覆盖、网络资源、网络性能、网络结构、网络稳定性等)采取相应措施进行解决。

通过以上投诉处理流程,用户投诉可以得到被动的事后解决。但这样的投诉处理模式停留在投诉的事后处理上,缺乏投诉预判和事前防范,问题的解决有不可控的时延,无法保障用户体验;与此同时也产生了较高的投诉处理成本。

针对以上问题,本文利用大数据及机器学习算法,挖掘投诉相关的网络性能指标,建立网络投诉预测模型及系统,对潜在的客户投诉进行预判。旨在协助运营商改善服务质量,系统化和自动化地提高投诉预判及防范能力,有效提升用户感知及自身竞争力。

2 基于决策树的高投诉小区预判方法

机器学习在网络优化和运维工作中得到了广泛的应用^[3-5]。分类决策树属于监督学习,可以根据特征值递归得到,将输入空间(即特征空间)划分为有限个类别。它的优点在于:一、计算复杂度低,且模型容易

可视化;二、算法完全不受数据缩放的影响,由于每个特征单独处理,且数据划分并不依赖缩放,因此决策树无需特征预处理。与此同时,当尺度不同的特征同时存在时,不影响决策树的效果。缺点在于可能会出现过拟合,泛化性较差。针对这个缺点可以通过剪枝或者组合树来对决策树进行优化。

本方案将LTE KPI作为特征集,将高投诉小区作为分类标签,构建CART(Classification And Regression Tree)分类决策树。具体步骤如图1所示。

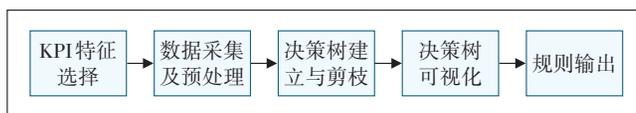


图1 基于决策树的高投诉小区预判流程

2.1 KPI特征选择

LTE KPI数据作为衡量4G网络质量、性能和业务质量的标杆,是网络监控、分析、评估与优化工作的重要指标^[6]。KPI能够一定程度上量化网络质量及用户感知,反应端到端的网络质量,在本质上与用户投诉中的问题有很强的相关度。合理地选择KPI作为决策树模型的特征集合,有助于更精准地构建基于网络性能的投诉预测模型。LTE性能指标主要涉及接入性、保持性、移动性、完整性、资源负荷等五大类关键指标。经过对KPI指标进行特征筛选,本方案中选取9个KPI指标作为决策树的特征指标,如表1所示。

2.2 数据采集及预处理

2.2.1 特征数据集

采集特定地(市)的LTE KPI数据并入库,并以省份、地(市)、小区、天、小时为维度对KPI完成数据计算,然后筛选上述9项KPI作为特征指标。为了提高预测模型的准确性,输入数据样本的准确性和有效性显得非常重要,需对数据进行清洗和预处理。

2.2.2 标签数据

从电子运维系统中采集与KPI对应日期的投诉数据,筛选出网络制式为4G,投诉类型为通信质量大类的投诉记录,并根据相同的维度(省份、地(市)、小区、天、小时)统计投诉数。对于评估所在的地(市)、日期范围内,第d天第h小时c小区的用户数 $U_{d,h,c}$,因通信质量原因的投诉数记为 $C_{d,h,c}$ 。平均用户投诉数可表示为:

$$P_{d,h,c} = C_{d,h,c} / U_{d,h,c} \quad (1)$$

将“是否为高投诉小区”作为标签数据,在此根据 $P_{d,h,c}$ 占比判断是否为高投诉小区,其中 T_0 为判决门

表1 KPI特征集

序号	KPI	指标计算
1	RRC连接建立成功率	RRC连接建立成功率 = RRC连接建立完成次数 / RRC连接请求次数(不包括重发) × 100%
2	E-RAB建立成功率	E-RAB建立成功率 = E-RAB建立成功总次数 / E-RAB请求建立总次数 × 100%
3	eNB内切换成功率	eNB内切换成功率 = (eNB内切换出成功次数 - 通过重建回源小区的eNB内切换出执行成功次数) / (eNB内切换出尝试次数) × 100%
4	eNB间切换成功率	eNB间切换成功率 = (eNB间切换出成功次数 - 通过重建回源小区的eNB间切换出执行成功次数) / (eNB间切换出尝试次数) × 100%
5	寻呼拥塞率	寻呼拥塞率 = 寻呼记录发送不成功次数 / 寻呼记录应该发送次数 × 100%
6	E-RAB掉话率	E-RAB掉话率 = (eNB触发的E-RAB异常释放总次数 + 小区切换出E-RAB异常释放总次数) / E-RAB建立成功总次数 × 100%
7	无线掉线率	无线掉线率 = (eNB发起的S1 RESET导致的UE Context释放次数 + UE Context异常释放次数) / UE Context建立成功总次数 × 100%
8	下行PDCP SDU丢包率	下行PDCP层丢包率 = 小区用户面下行PDCP层丢包数 / 总发包数 × 100%
9	下行PDCP层平均速率	PDCP层下行平均速率 = 下行PDCP层下行速率的均值

$$Flag_{d,h,c} = \begin{cases} 0, & P_{d,h,c} \leq T_0 \\ 1, & P_{d,h,c} > T_0 \end{cases} \quad (2)$$

经过上述步骤,特征数据集和标签数据集整理如表2所示。

表2 数据集

序号	类型	指标	序号	类型	指标
1	特征	RRC连接建立成功率	6	特征	E-RAB掉话率
2	特征	E-RAB建立成功率	7	特征	无线掉线率
3	特征	eNB内切换成功率	8	特征	下行PDCP SDU丢包率
4	特征	eNB间切换成功率	9	特征	下行PDCP层平均速率
5	特征	寻呼拥塞率	10	标签	高投诉小区标识

2.3 决策树建立与剪枝

在表2中选取10 000个数据样本,拆分为训练集与测试集,使用整个数据集的80%作为训练集,20%作为测试集。使用训练集对决策树进行训练,而后将测试集作为输入得到预测的标签集合。随后针对本文的二分类问题,通过混淆矩阵的方式分别统计分类模型归错类、归对类的观测值个数,评估模型最终的效果。最后使用graphviz将决策树可视化(见图2)。

为了防止决策树过拟合,这里采用预剪枝的方法,其中Flag_{d,h,c} = 1记为高投诉小区。

限,其中Flag_{d,h,c} = 1记为高投诉小区。

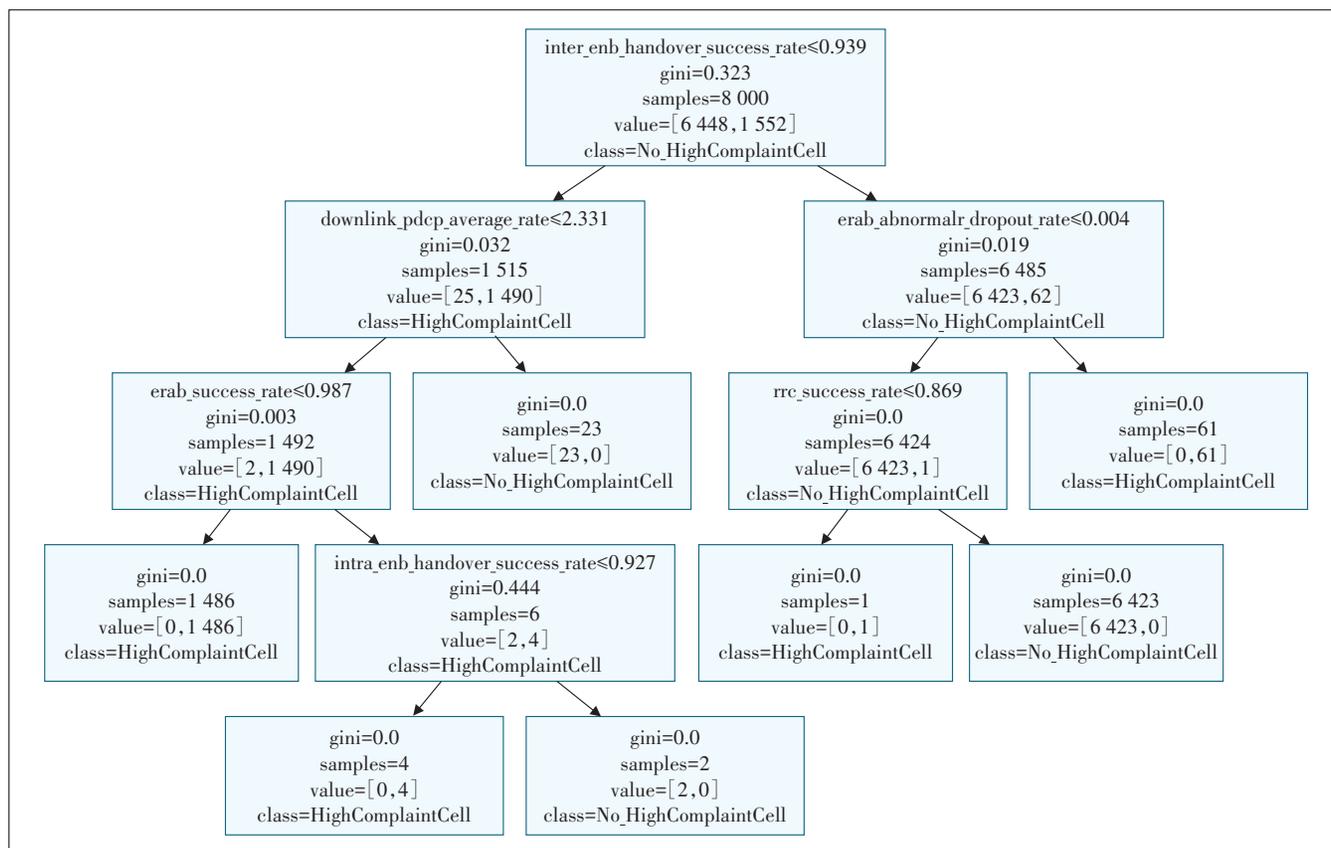


图2 决策树可视化

法。预剪枝的限制条件包括限制树的最大深度、限制叶节点的最大数目,或者规定一个结点中数据点的最大

小数目来防止继续划分等,这些在DecisionTreeClassifier的参数中进行设置(见表3)。

表3 模型数据

分类	eNB间切换成功率	下行PDCP层平均速率	E-RAB掉话率	E-RAB建立成功率	RRC连接建立成功率	eNB内切换成功率
高投诉小区	小于0.939	小于2.331	-	小于0.987	-	-
	小于0.939	小于2.331	-	大于0.987	-	小于0.927
	大于0.939	-	小于0.004	-	小于0.869	-
	大于0.939	-	大于0.004	-	-	-
非高投诉小区	小于0.939	大于2.331	-	-	-	-
	小于0.939	小于2.331	-	大于0.987	-	大于0.927
	大于0.939	-	小于0.004	-	大于0.869	-

3 方案部署及应用

从具体实施的角度,本方案部署在网络中还需要在运营商电子运维系统基础上部署2个关键的子系统:数据采集及处理子系统和策略判决及执行子系统。具体的方案如下。

数据采集及处理子系统,对于网络KPI进行实时采集和处理,按照指定维度对各特征KPI进行数据输

出。数据进入策略判决及执行子系统后,该子系统会依据现存的静态策略,对各小区是否为疑似高投诉小区进行预判,如果该小区满足高投诉小区的特征,则将其纳入高投诉小区名单。最终策略判决及执行子系统将生成的高投诉小区名单输出给电子运维平台或者网优平台,由相关网优或者运维人员对目标小区进行问题定位及解决(见图3)。

与此同时,通过结合当前模型的预测和实际网络

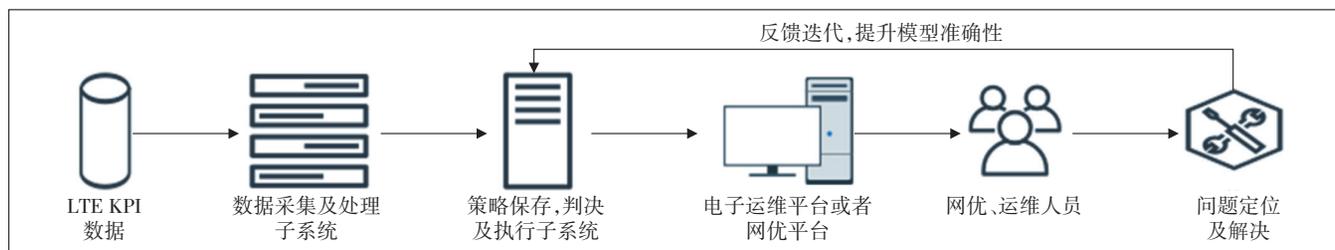


图3 方案部署及应用

问题,对模型进行反馈调节,通过迭代提高预报模型的效率和精度。

4 总结

本文构建了基于网络性能的投诉预测模型,对潜在的客户投诉进行预测。建立一套打通KPI与前端投诉数据的分析体系,统筹分析网络的用户感知数据,使前后端通过数据的分析能够联动起来,通过数据的联合分析结果,对市场、客服、建设、维护多个方面起到支撑作用。改善投诉处理业务流程,降低企业投诉处理成本,提高运营商的服务质量和竞争力。

参考文献:

[1] 崔小丽,贾金良,王慧慧. 基于大数据的投诉关联分析及平台构建[J]. 通信管理与技术,2016(3):74-77.
 [2] 董智纯,杨林,詹念武. 一种基于大数据技术的投诉分析与预测

系统[J]. 信息通信,2015(9):291-292.
 [3] 王希. 基于MR数据与机器学习的LTE用户感知评估方法[J]. 移动通信,2018(8):26-31.
 [4] 钱兵,王兵. 基于机器学习算法的LTE小区智能化分类[J]. 电信技术,2018(5):16-19.
 [5] 王磊. 基于机器学习技术的LTE网络智能优化系统设计[J]. 电信工程技术与标准化,2018(1):44-47.
 [6] 李峻洋,赵占强,郭省力. LTE无线网络优化关键性能指标研究[J]. 邮电设计技术,2014(4):83-86.

作者简介:

曹丽娟,毕业于北京邮电大学,工程师,硕士,主要从事数据分析及可视化、运营商大数据挖掘及行业应用工作;程新洲,教授级高级工程师,主要从事大数据研究与应用工作;徐乐西,毕业于伦敦大学,高级工程师,博士,主要从事大数据分析与应用、LTE-A/5G移动通信研究及标准化工作;张涛,毕业于北京邮电大学,工程师,硕士,主要从事运营商大数据挖掘及行业创新应用产品的研究工作;贾玉玮,工程师,硕士,主要从事机器学习、大数据行业应用、大数据移动网络分析等工作;成晨,毕业于北京邮电大学,工程师,硕士,主要从事通信大数据挖掘等技术领域的研究工作。