

面向F5G的光接入网智能规划

Research on Intelligent Network Planning
of Optical Access Network for F5G

应用研究

李奥, 韩赛, 王光全, 邵岩, 王泽林(中国联通研究院, 北京 100048)

Li Ao, Han Sai, Wang Guangquan, Shao Yan, Wang Zelin(China Unicom Research Institute, Beijing 100048, China)

摘要:

针对在保障用户带宽需求的前提下,如何提前评估扩容升级并主动规划光接入网络的问题,提出了一种基于树模型的接入网扩容规划算法,即通过决策树ID3算法定量分析影响PON口峰值速率的关键因素,根据决策树的结果使用回归树CART算法预测PON口峰值速率区间。实验表明,基于该方案构建的2个模型准确率均达80%以上,可为F5G时代光接入网精准评估并规划扩容提供借鉴。

关键词:

人工智能; F5G网络; 光接入网; 扩容评估

doi:10.12045/j.issn.1007-3043.2023.02.013

文章编号:1007-3043(2023)02-0070-04

中图分类号:TN915.6

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

Aiming at the problem of how to evaluate the capacity expansion and upgrade in advance and actively plan the optical access network on the premise of ensuring the bandwidth demand of users, an optical access network capacity expansion planning algorithm based on the tree model is proposed. The key factors affecting the peak rate of PON ports are quantitatively analyzed through the decision tree algorithm (ID3). The peak rate range of the PON ports is predicted by the regression tree algorithm (CART) according to the result of a decision tree. Experiments show that the accuracy of the two models based on the scheme is more than 80%, which can provide a reference for the accurate evaluation and planning expansion of optical access network in the F5G era.

Keywords:

Artificial intelligence; F5G network; Optical access network; Capacity expansion evaluation

引用格式:李奥,韩赛,王光全,等.面向F5G的光接入网智能规划应用研究[J].邮电设计技术,2023(2):70-73.

1 概述

2021年的《政府工作报告》中明确指出,要加大5G网络和千兆光网建设力度,丰富应用场景。这是“千兆光网”首次被写入政府工作报告,F5G作为千兆光网建设的第5代固定网络技术,成为我国重要的新型基础设施^[1]。随着我国加快推进千兆光网建设,千兆光网的发展势如破竹,固定网络已经进入F5G时代^[2]。

截至2021年底,三大运营商的固定互联网宽带接入用户总数达5.36亿户,其中1千兆及以上接入速率

的用户数达到3 456万户^[3],根据工信部发布的《“双千兆”网络协同发展行动计划》以及国务院常务会议要求,预计到2023年底,千兆光纤网络具备覆盖4亿户家庭的能力^[4]。

为配合千兆宽带上线,运营商需将原有的1G PON扩容到10G PON。然而,评估规划接入设备升级扩容不是一蹴而就的,而且在运营商投资成本受限时,扩容工作也需要分阶段持续进行。当前主要通过判断PON口下是否具有千兆用户,或者根据PON端口单位时间峰值流量(即峰值速率)拥塞后上报的工单作为评估扩容的依据。当前方式主要依赖人工经验判断,无法量化PON口峰值速率影响因子,特别是在运营商

收稿日期:2022-12-22

成本受限的情况下, 依靠经验难以更精准地找到有升级需求的节点。另外, 收到工单后由人工重新评估规划, 会导致处理时间变长, 扩容不及时, 影响用户体验。如何量化 PON 口峰值速率影响因子, 预测 PON 口峰值速率区间, 更合理地提前评估规划设备升级扩容是当前亟需解决的问题之一。因此需要引入智能化的手段来解决上述问题。

随着 AI 的蓬勃发展, AI 技术与 F5G 的深度融合将助力解决 F5G 网络新问题, 驱动智慧光网络不断升级^[5]。目前扩容规划领域的扩容规划算法主要有: 基于相似度分析的扩容规划算法^[6]、基于网格化的扩容规划算法^[7]、基于用户特征分析的扩容规划算法^[8]、基于最小生成树的规划算法^[9-10]等, 但尚未有针对 PON 口峰值速率影响因子量化, 构建 PON 口全参量模型的研究。而监控 PON 口流量(OLT PON 口利用率), 可以避免端口出现拥塞, 影响端口下挂用户体验; 提前预判拥塞迹象, 为 PON 口下用户的发展和迁移及 PON 口扩容(1G PON 到 10G PON)提供指导; 通过流量异常的 PON 口挖掘出流量异常的用户, 辅助市场部的用户营销工作。

针对以上问题, 本文提出了一种面向 F5G 光接入网的基于决策树和回归树算法的扩容规划算法, 可实现全面综合的量化 PON 口峰值速率影响因子, 预测 PON 口峰值速率区间, 精确指导扩容升级, 克服人工方案的局限性。该算法采用决策树(ID3)算法^[11]对 PON 口峰值速率影响因子进行量化分析, 可有效量化扩容指标; 使用回归树(CART)算法^[12]预测 PON 口峰值速率区间, 为精准规划扩容提供科学依据。为进一步优化实验效果, 本文采用 SMOTE 均衡化函数解决样本类数量不均衡导致的算法结果具有倾向性的问题, 并通过实验测试验证了本算法的效果。

2 基于决策树和回归树的扩容规划算法

2.1 数据

本文采用的数据是从某运营商某市现网获取的, 约有 290 万条号线数据、1 700 万条 AAA 数据、15 万条 PON 口数据和 290 万条经分数据, 经过数据关联统计后作为本文分析计算的基础。经过处理后, 每个 PON 口的特征包括数据采集时间、OLT 的 ip 号、OLT 的端口号、各 PON 口值(合并 OLT 的 ip 号和端口号)、流出峰值速率、流入峰值速率、用户套餐速率、PON 口下挂用户的均值带宽、PON 口下挂用户的均值经纬度、PON

口下挂用户的均值在线时长、PON 口下 IPTV 用户数、PON 口下用户数总数、PON 口下公众客户/商务客户/大客户人数。

2.2 扩容规划算法

本文基于决策树算法和回归树算法提出扩容规划算法。决策树(ID3)算法是一种经典的分类算法。该算法主要依据特定属性划分后数据的信息增益, 依次选择具有最大信息增益的特征作为划分依据进行分类; 回归树(CART)算法是一种经典的回归算法, 采用平方误差最小化准则作为节点划分的依据, 将预测区间范围内的数据均值作为预测结果, 实现非线性回归预测。如图 1 所示, 具体流程主要包括用决策树算法对 PON 口峰值速率影响因子进行定量分析和使用回归树算法预测 PON 口峰值速率区间 2 个步骤。

a) 采用决策树算法对 PON 口峰值速率影响因子进行量化分析。

(a) 输入数据和参数, 设置决策树算法的树深度 max_depth=100, 随机种子设为 0, 随机选择 80% 作为训练集数据, 20% 作为测试集数据。

(b) 对输入的各 PON 口数据打标签(训练集和测

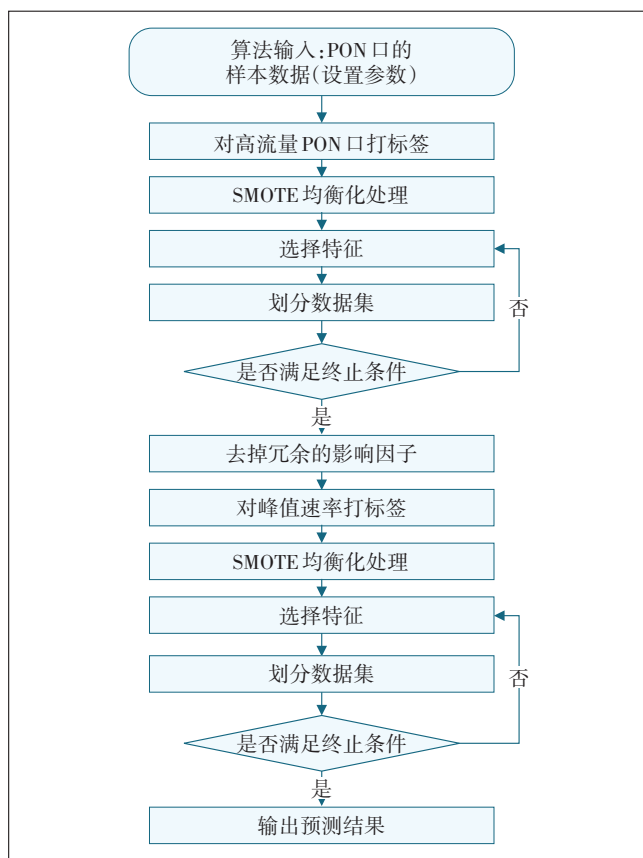


图 1 扩容规划算法实现流程图

试集都打标签),高流量 PON 口记为 1,其他 PON 口记为 0。高流量 PON 口是指高流量用户所在的 PON 口,通过统计全网每个用户单日累计的流量数值,选取前 10% 的用户记为高流量用户。

(c) 调用 imblearn 包中的 SMOTE 函数对步骤(b)处理后的数据做均衡化处理,首先对每一个小样本类样本 (x, y) 按欧氏距离找出 K 个最近邻样本,再从 K 个最近邻样本中随机选择 N 个样本进行随机线性插值,直到大、小样本数量均衡。其中指定近邻数设置为 5,从近邻样本中随机挑选的样本个数设置为 10,过采样所需 CPU 数量设置为 1。

(d) 计算各属性的信息熵和信息增益,选择信息增益最大的属性作为当前的特征。信息熵和信息增益的计算公式为:

$$\text{Entropy}(X) = -\sum_{i=1}^m p(u_i) \log_2 p(u_i) \quad (1)$$

其中, $p(u_i) = \frac{|u_i|}{|S|}$, $p(u_i)$ 为类别 u_i 在样本 S 中的概率。

$$\text{IG}(S, A) = \text{Entropy}(S) - \sum_{V \in \text{value}(A)} \frac{|S_V|}{S} \text{Entropy}(S_V) \quad (2)$$

其中, A 表示样本属性, $\text{value}(A)$ 是属性 A 所有取值的集合。 V 是 A 的其中一个属性值, S_V 是 S 中 A 的值为 V 的样例。

(e) 将当前的特征作为数据集划分的依据。

(f) 如果没有属性可以再分了或者划分出来的类已经属于同一类了,则本循环结束,利用 `tree.export_graphviz` 函数输出 PON 口峰值速率影响因子数值;如果还有属性可分则返回步骤(d)。

b) 根据 PON 口峰值速率影响因子量化结果,使用回归树算法预测 PON 口峰值速率区间,过程如下:

(a) 对步骤 a) 中步骤(f)的计算结果从大到小排序,去掉最小的影响因子(数据采集时间)。

(b) 对所有数据中的流入流出峰值速率画散点图,确定不同流速对应的 PON 口数分布情况,并根据该分布情况为不同区间打标签(训练集和测试集都打标签),尽量缩小区之间的 PON 口总数差距,区间值与标签的对应关系见表 1。

(c) 调用 imblearn 包中的 SMOTE 函数对步骤(b)处理后的数据做均衡化处理,该方法原理与步骤 a) 中的步骤(c)相同。

(d) 依次遍历每一个特征 j , 以及该特征的每一个

表 1 PON 口流入流出峰值流速区间及其对应的标签值

流入峰值标签	流入峰值/(Mbit/s)	流出峰值标签	流出峰值/(Mbit/s)
0	0~20	0	0~30
1	20~30	1	30~100
2	30~40	2	100~120
3	40~50	3	120~150
4	50~60	4	150~180
5	60~70	5	180~200
6	70~100	6	200~400
7	100~150	7	400~600
8	150~500	8	600~800
-	-	9	800~1 200

取值 s , 计算每一个切分点的损失函数。选择损失函数最小值作为切分点。损失函数最小值 RL_{\min} 计算公式为:

$$R_1(j, s) = \{x | x_j < s\} \quad (3)$$

$$R_2(j, s) = \{x | x_j > s\} \quad (4)$$

$$RL_{\min} = \min \left\{ \sum_{x_i \in R_1(j, s)} (y_i - y_{R_1, \text{mean}})^2 + \sum_{x_i \in R_2(j, s)} (y_i - y_{R_2, \text{mean}})^2 \right\} \quad (5)$$

其中, y 为输出的连续变量,将输入划分为 2 个区域,分别是 R_1 和 R_2 , $y_{R_1, \text{mean}}$ 与 $y_{R_2, \text{mean}}$ 分别为每个区域内输出的平均值。

(e) 使用步骤(d)计算出来的切分点将当前的输入空间划分为 2 个部分。

(f) 如果没有特征可以继续划分了,则本循环结束;如果还有特征可以分则返回步骤(d)对划分后的 2 个部分再重新计算切分点。

(g) 输出预测的 PON 口峰值速率区间。

2.3 实验结果

根据以上方法获得的 PON 口峰值速率影响因素如表 2 所示。从整体上来看,大客户人数和数据采集时间对 PON 口流入流出峰值速率均影响较小。

根据以上方法获得的 PON 口峰值速率区间预测值如表 3 所示。由于 PON 口数量较多,本文只展示部分实验结果。

PON 口峰值速率影响因子量化准确率及 PON 口流入流出峰值速率区间预测准确率如表 4 所示。

本实验采用 python 语言,对 PON 口峰值速率影响因素进行了综合的量化分析,实现了对 PON 口的流入流出峰值速率区间的预测。根据实际观察,当前 PON 口流量并未达到阈值,节假日流量较平时略有增加,

表2 PON口峰值速率影响因子量化值

PON口影响因子	流入峰值	流出峰值
公众客户人数	0.248	0.398
PON口所带用户均值经度	0.09	0.095
PON口所带用户均值纬度	0.104	0.109
是IPTV人数占比	0.07	0.060
用户套餐速率	0.0371	0.032
均值在线时长	0.055	0.054
商务客户人数	0.273	0.128
均值带宽	0.092	0.101
大客户人数	0.023	0.015
数据采集时间	0.008	0.008

表3 PON口峰值速率区间预测结果(节选)

峰值速率区间实际值	峰值速率区间预测值
4.0	7.0
3.0	2.0
6.0	6.0
0.0	1.0
9.0	8.0
10.0	10.0
1.0	1.0

表4 PON口峰值速率影响因子量化和区间预测值准确率

峰值流速	影响因子量化准确率/%	区间预测值准确率/%
流入	95.0	82.7
流出	94.8	80.9

依旧未达到扩容阈值。节假日不定期、不显著的流量增加不影响运营商扩容逻辑。因此本实验中未引入节假日等因素。另外, PON口峰值速率受数据采集时间因素的影响较少,影响PON口峰值速率的因素是多方位的,其中用户类型和套餐类型等因素都需要考虑在内,且PON口峰值速率本身具有突发性,通过学习PON口峰值速率随时间变化的趋势来预测未来一段时间PON口峰值速率的意义不大,因此,本实验是通过向本方法构建的模型中输入未来一段时间内这些影响因子最可能的数值,来预测该段时间PON口可能的峰值速率。本实验改进了传统依赖人工经验判断的方式,通过该案例的计算方法,可以更准确地获取PON口峰值速率影响因素量化值,减少了人的主观影响和计算中弱相关参数造成的冗余。

3 结束语

本文研究如何在 F5G 网络分阶段上线千兆带宽

的要求下,量化PON口峰值速率影响因子,预测PON口峰值速率,有效提高PON口设备规划扩容的精准性,结合AI技术提出一种基于决策树和回归树的光接入网扩容规划算法。通过决策树算法定量分析影响PON口峰值速率的关键因素,根据决策树的结果,使用回归树算法预测PON口峰值速率区间,最终实现PON口峰值速率的量化,为PON口设备规划扩容提供更全面的参考依据。实验结果表明,该算法可以量化PON口峰值速率影响因子,提高预测PON口峰值速率区间的准确率。

参考文献:

- [1] 海峰看科技. 如何构建城市智能体“神经系统”? F5G城市光网或是最优选择[EB/OL].[2022-08-02]. <https://baijiahao.baidu.com/s?id=1740033640031799067&wfr=spider&for=pc>.
- [2] 李鑫,赵永利,李卓桐,等. 第5代固定通信网F5G发展展望[J]. 通信世界,2021(4):32-36.
- [3] 佚名. 行业保持稳中向好运行态势 电信业务收入比上年增长8.0%[N]. 人民邮电,2022-01-27(3).
- [4] 佚名. 工信部印发《“双千兆”网络协同发展行动计划(2021-2023年)》[J]. 互联网天地,2021(4):3-11.
- [5] 吴军. 智慧光网络为千行百业数字化转型注智赋能[N]. 人民邮电,2021-09-28(6).
- [6] 占天顺. 基于相似度分析的接入网网络规划算法研究[D]. 北京:北京邮电大学,2021.
- [7] 赵涛. 基于网格化的宽带接入网规划方法浅析[J]. 通讯世界,2019,26(7):177-178.
- [8] 孙静蕾. 基于用户特征分析的接入网规划方法的设计与实现[D]. 北京:北京邮电大学,2020.
- [9] 孙利浩,张晓炎,陈思超,等. 基于最小生成树的配电通信接入网规划算法[J]. 电力学报,2019,34(4):336-343.
- [10] YAN J, LU S W, WANG S C, et al. A multi-level investment allocation indicator system for distribution network planning[J]. Frontiers in Energy Research, 2022, 10:889325.
- [11] ZHANG T S, MCMILLAN W W. An implementation of ID3 that creates executable rules (abstract) [C]//Proceedings of the 1990 ACM annual conference on Cooperation. Washington, D. C., USA: Association for Computing Machinery, 1990:450.
- [12] YE H C H. Classification and regression trees (CART)[J]. Chemometrics and Intelligent Laboratory Systems, 1991, 12(1):95-96.

作者简介:

李奥,助理工程师,硕士,主要从事网络智能化研究工作;韩赛,高级工程师,博士,主要从事网络智能化研究工作;王光全,教授级高级工程师,硕士,主要从事光通信、量子加密通信等领域研究工作;邵岩,高级工程师,硕士,主要从事光网络相关技术和承载业务研究工作;王泽林,高级工程师,硕士,主要从事IP技术研究、IP网络与设备研究及应用相关工作。