

# 构建基于AI算法的无线网络智能 扩容优化体系

## Construction of Radio Network Intelligent Expansion Based on AI Algorithm

陈 锋,许绍松,李张铮(中国联通福建省分公司,福建 福州 350000)

Chen Feng,Xu Shaosong,Li Zhangzheng(China Unicom Fujian Branch,Fuzhou 350000,China)

### 摘 要:

针对无线网络面临的扩容投资受限、扩容不及时、扩容标准单一等问题,通过引入KMeans聚类算法、KK关联规则等AI智能算法,实现小区场景聚类、流量压抑识别与测算,并通过自动化工具开发实现智能优化和自动输出扩容方案。研究表明,该模型可智能识别高价值流量压抑小区并估算出压抑流量,可作为精准智能扩容的基础平台。

### 关键词:

流量压抑识别;KMeans聚类;KK关联规则;网络优化;网络扩容

doi:10.12045/j.issn.1007-3043.2020.10.008

文章编号:1007-3043(2020)10-0036-05

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:

Aiming at the problems faced by radio network such as limited expansion investment, untimely expansion and single expansion standard, the clustering of cell scenes, the identification and measurement of traffic suppression are realized by introducing the AI intelligent algorithm such as KMeans clustering algorithm and KK association rules, and the intelligent optimization and automatic output expansion scheme are realized through the development of automation tools. Research results show that the model can intelligently identify high-value traffic suppression cells and estimate suppression traffic, which can be used as the basic platform for precision and intelligent expansion.

### Keywords:

Traffic suppression recognition; KMeans clustering; KK association rules; Network optimization; Network expansion

引用格式:陈锋,许绍松,李张铮. 构建基于AI算法的无线网络智能扩容优化体系[J]. 邮电设计技术, 2020(10): 36-40.

## 0 引言

近年来,2I2C业务的迅速发展以及不限量套餐的推出,带来数据业务流量的爆发式增长,给无线网络带来巨大压力,直接影响到用户感知,而扩容投资成本受限,低价值用户、流量、场景区域扩容导致流量效益下降。如何在保障用户感知的同时,降低投资成本,提升扩容效能,成为摆在网优工作面前的一个紧迫课题。动态合理地识别高负荷小区,很大程度上依赖于流量压抑小区的识别,如果能够智能识别小区流量的压抑状态,那么就可及时做出相应的优化扩容方

案安排,提高资源利用率。

机器学习技术作为人工智能的重要组成部分,是国家发展战略重点扶持的目标<sup>[1]</sup>,是当下各行业关注应用的焦点。为了推动日常高负荷小区优化扩容的智能化,提升网络运营智能化水平,有必要研究基于机器学习算法的流量压抑小区识别。

## 1 小区传统扩容方法面临的痛点

### 1.1 缺乏价值评估体系,扩容投资与小区场景价值脱钩

由于用户日益增长的网络质量及网络发展需求,网络运营压力不断增大,但由于网络发展投资有限且在网络建设中采取用户感知容量保障方案,高价值用户及高价值场景不能得到较好的网络质量,导致网络

收稿日期:2020-09-11

发展效率较低,不能实现投资效益最大化。

## 1.2 扩容周期长,扩容存在滞后性,影响用户感知

日常扩容采取的方法为前期建设后期评估,虽然可以用优化手段进行多维度分析,但是扩容工作为后发型,即发现存在容量问题或者用户感知较差问题后进行站点软扩或补点扩容,从扩容申请到设备下发及扩容完成周期较长,影响用户体验,而且还会因为人员流动等原因,造成扩容后该区域已不是热点区域,扩容达不到理想效果,用户感知较差。

## 1.3 采用统一扩容标准,评价指标一刀切导致流量释放效果不佳

传统的扩容方法中高价值场景分类及扩容标准一刀切,没有考虑不同场景的业务模型不同、网络质量不同导致的每个小区流量压抑的临界点不一样,采取同一标准会出现无需扩容误扩容和需要扩容未扩容的情况。

## 2 基于聚类算法的小区流量压抑识别

导致流量压抑的因素很多,不同的社会场景、不同的人数规模、不同的无线环境、不同时段、不同业务类型、不同用户行为等都会导致流量压抑。如何综合考虑这些因素,预估出被压抑的流量已成为无线网络运营商的业务痛点之一。

本文通过研究聚类算法、线性回归、关联规则等机器学习算法,用分类、拟合、压抑识别、压抑预估4个步骤完成了对每个小区压抑流量预估,使得运营商在新建站、扩容站时可以根据实际的用户需求完成站点的规划及扩容。

### 2.1 基于KMeans算法的小区场景聚类

#### 2.1.1 KMeans聚类算法简述

所谓聚类算法是指将一堆没有标签的数据自动划分成几类的方法,属于无监督学习方法,这个方法要保证同一类的数据有相似的特征。根据样本之间的距离或者说是相似性(亲疏性),把越相似、差异越小的样本聚成一类(簇),最后形成多个簇,使同一个簇内部的样本相似度高,不同簇之间差异性高。

KMeans算法步骤如下。

- 随机选取 $k$ 个样本点作为初始起点。
- 计算所有点到初始点距离,并将距离最近的点集合作为同类。
- 重新计算同类的重心(采用欧氏距离)。
- 重新计算所有点到新的分类中心的距离,以重

心为新的起点。

- 回到步骤b),重新划分分类和重心。

KMeans算法特点如下。

- 每一类趋向球形(因为距离定义是欧氏距离)。
- 分类受到初始中心点的选择,通常多次运行,取较好的结果。
- 算法的时间复杂度为 $O(tkmn)$ ,其中 $t$ 为迭代次数, $k$ 为聚类簇个数, $m$ 为特征数, $n$ 为样本数。一般 $t$ 、 $k$ 、 $m$ 均可认为是常量,所以时间复杂度可以简化为 $O(n)$ 即呈线性,故算法运行速度很快。

- 算法属于启发式算法,无法保证收敛到全局最优解,但在大多数情况下,可以收敛到比较理想的局部解。

无线小区间特征存在一定的相似性,比如用户数、忙闲时、大中小包业务等,这就使得可以通过计算小区特征间的距离来衡量小区是否属于同一类场景,而且人工对小区打场景标签存在主观性和盲目性,不适合选择有监督的分类算法对场景进行建模,所以本文选择无监督的KMeans聚类算法来对小区场景进行自动聚类。

#### 2.1.2 基于KMeans算法的无线小区聚类

传统场景划分算法大致分3类:根据基站所处位置划分场景(地理场景),根据基站覆盖类型、服务对象划分场景(服务场景),根据基站天线角度、功率、站高等信息划分场景(无线场景)。

由于城区无线基站数量异常庞大,而且通信业务已由原先单一的语音和短信方式发展到现在多样化的数据业务(视频、微信、浏览),再按照原先的方式划分已经不能实现无线场景的准确划分。另外,面对众多的流量压抑需扩容小区,在投资有限的情况下,采用统一用户感知容量保障策略,高价值用户无法得到更好的用户体验,急需在小区场景中引入用户价值属性。因此,基于AI的无线网络智能扩容提出一种新的场景划分方法,从真实数据角度出发,通过应用机器学习算法中的聚类算法,输出更精确的小区场景模型。

聚类算法输入的训练数据特征主要包括用户数、流量、忙闲时、大中小包业务、场景、频段、带宽、用户出账收入、用户套餐等,采集每小时每小区数据生成训练样本。

- 频段、场景识别:按照地理位置识别出场景(商场、学校、道路、住宅区、车站等),在同一场景下,按照

不同频段不同带宽进行分组。

b) 时段学习:学习每个场景的忙闲时(输入参数平均RRC连接数、PRB利用率等)数据。

c) 大、中、小包分组:根据每ERB流量按照聚类区分大、中、小包3类采样点。

将样本数据输入KMeans算法进行训练,同时选择合适的聚类数量 $K$ ( $K$ 值的选择在2.1.3条中描述)。

### 2.1.3 小区聚类效果评价

KMeans聚类最佳聚类数目 $K$ 值的选择主要有手肘法和轮廓系数法2种<sup>[2]</sup>。

手肘法:手肘法根据误差平方和(SSE——sum of the squared errors)衡量聚类数目 $K$ 的合理性,其定义如下:

$$SSE = \sum_{p \in C_i} |p - m_i|^2 \quad (1)$$

式中:

$C_i$ ——第 $i$ 个簇

$P$ ——簇 $C_i$ 里的样本点

$m_i$ ——簇 $i$ 的质心

手肘法的核心思想是:随着 $K$ 值的逐步增大,各样本趋于合理聚类,把越相似、差异越小的样本聚成一类,每个类内的样本误差平方和逐步降低。当 $K$ 小于最佳值时,样本间不断重组聚类,SSE曲线下降幅度较大;当 $K$ 趋于最佳值时,样本重组趋于稳定,SSE曲线下降幅度骤减;当 $K$ 超过最佳值后,SSE曲线基本平缓。SSE和 $K$ 的关系图是一个手肘的形状,而这个肘部对应的 $K$ 值就是算法的最佳聚类数,这也是该方法被称为手肘法的原因。

轮廓系数(Silhouette Coefficient)法:某个样本点 $x_i$ 的轮廓系数定义如下:

$$S = \frac{b - a}{\max(a, b)} \quad (2)$$

式中:

$a$ ——凝聚度,是样本点 $x_i$ 和同簇其他样本的平均距离

$b$ ——分离度,是样本点 $x_i$ 和距离最近簇中其他样本的平均距离最近簇定义如下:

$$C_i = \arg \min_{C_k} \frac{1}{n} \sum_{p \in C_k} |p - x_i|^2 \quad (3)$$

其中 $p$ 是簇 $C_k$ 中的样本。轮廓系数范围在 $[-1, 1]$ , $S$ 接近1,则说明样本 $i$ 聚类合理;所有样本的 $S$ 的均值称为聚类结果的轮廓系数,该系数可度量该聚类是否合理、有效。

利用以上原理实现了不同 $K$ 值下的误差平方和轮廓系数的可视化呈现如图1所示,从图1中流量、用户数的聚类各种 $K$ 值的SSE曲线(蓝色)可以发现 $K=3$ 时曲率最高,同时 $K=3$ 时轮廓系数(红色点)为第2大值,根据肘分析法,本次聚类数目最佳值 $K$ 取3。

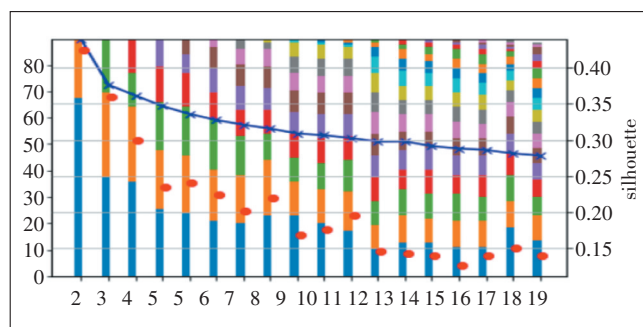


图1 根据手肘法和轮廓系数法选择最佳 $K$ 值

### 2.1.4 基于线性回归的场景标杆函数建立

使用线性回归的目的是建立场景级无压抑流量模型,本场景下的所有小区流量压抑情况可根据用户数等特征代入模型来进行计算。

对上一步得到的每个聚类场景,根据预先制定的专家优化规则,判定此时此刻此小区属于优质小区时刻或质差小区时刻,优质小区时刻总集称为标杆数据集。再根据标杆数据集,计算该场景下用户数和流量之间的关系模型即线性回归标杆函数,为增加鲁棒性,此处采用线性模型(见图2)。

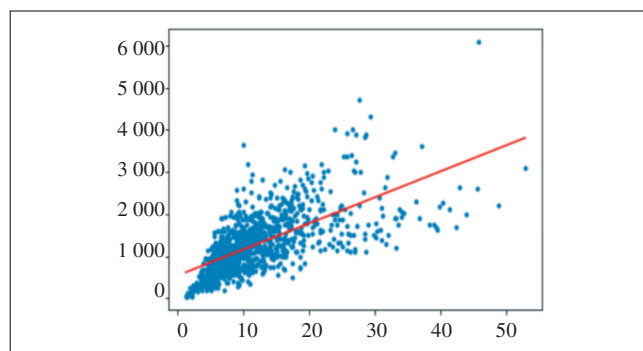


图2 流量和用户数线性标杆函数

## 2.2 流量压抑识别与测算

### 2.2.1 FPgrowth算法概述

最初提出关联规则的动机是购物篮分析问题。通过发现顾客放入“购物篮”中的不同商品之间的关联,分析顾客的购物习惯。这种关联的发现可以帮助零售商了解哪些商品频繁地被顾客同时购买,从而帮助他们制定更好的营销策略。

FPgrowth算法包含一个一棵FP树和一个项头表, 每个项通过一个结点链指向它在树中出现的位置。FP树的建立过程如下。

- 开始节点为空。
- 首先插入第1行关键字。
- 插入第2行关键字, 如果有重复的前缀路径, 则路径上的节点+1。
- 插入所有的数据之后, FP树和链表也都建好。

#### 2.2.2 基于FPgrowth算法的KPI-KQI关联规则

导致KQI劣化的原因可能不止是单个KPI异常, 可能是多个KPI异常, 引入机器学习的关联规则算法可以帮助定位引起KQI劣化的最大可能KPI原因。基于FPgrowth算法的小区KPI-KQI关联规则挖掘主要步骤如下。

- 取过去3~4个月内的质差小区, 去除非无线侧原因导致, 剩下无线质差小区的指标数据包括KPI、KQI和MR。
- 对指标数据进行处理, 构建FP树, 高效挖掘百万级的项集得到频繁项集及每个频繁项集的支持度。
- 计算频繁集的置信度、提升度, 输出关联指标规则库。

#### 2.2.3 流量压抑估算

由KK关联规则判断出无线原因导致的流量压抑小区后, 利用线性标杆函数算出该用户数应有总流量, 减去现有流量就是受压抑流量。针对这部分小区采取先优化后扩容的原则进行处理。

### 3 基于决策树的流量压抑小区优化

基于专家经验, 建立专家智能优化决策树, 对可优化站点自动识别, 从小区故障、流量突增、干扰导致的PRB虚高、邻区负荷等方面, 对目标小区进行检查, 对具备可优化特征的基站优先实施无线优化及与周边站点的话务均衡。

决策树的实现通过采集小区故障告警、小区工参、用户数、PRB利用率和流量、小区切换数据、CQI占比等数据, 自主开发优化工具, 自动定位网络问题, 并根据问题自动生成推荐的优化方案(见图3)。

### 4 基于决策树的智能扩容

针对无优化空间且优先级高的TOP小区根据专家经验结合真实情况构建小区智能扩容决策树, 输入扩容小区工参和指标自动匹配最佳扩容方案, 实现小

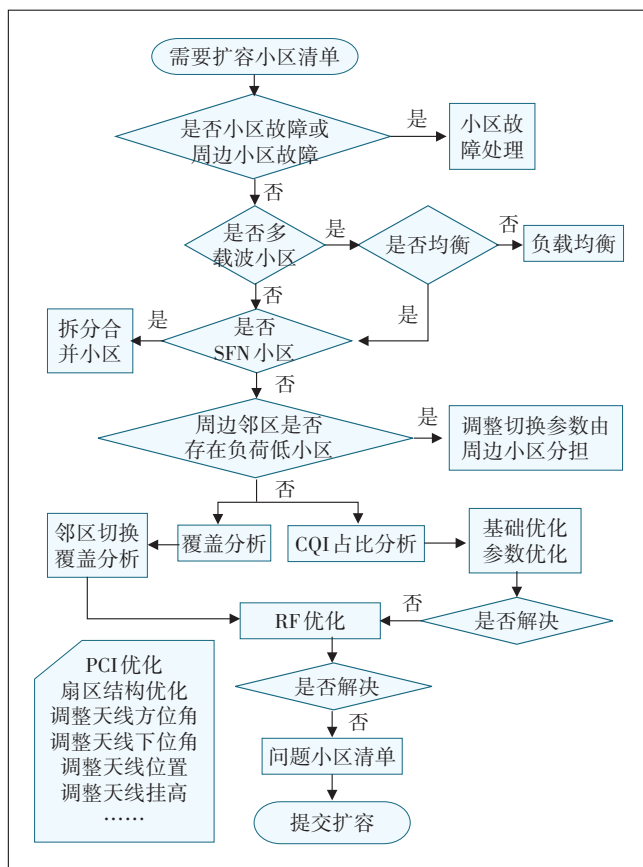


图3 智能优化决策树模型

区扩容方案智能输出。扩容措施主要有软件扩容、硬件扩容、小区劈裂、微站、室分等(见图4)。

- 根据专家经验设置扩容参数, 如用户数大于90/200、PRB大于70%, 需要进行L2100和L1800同时扩容等。

- 对BBU下基带板能力进行识别, 对于不同的基站不同的板件能力进行自动分析, 得出基站需要新增或更换板件。

- 根据小区需求的不同指定特有的扩容方式, 实现小区扩容后问题解决率提升及投资效益最大化。

- 采取闭环处理, 小区扩容后跟踪评估, 精准评估扩容效能。

### 5 基于AI的无线网络智能扩容研究的应用

2019年福州联通在现网选取高流量场景小区2 955个(含高校2 460个小区)作为样本进行试验。先将高校小区进行聚类, 成功地将这些小区分为八大场景类, 比如白天忙时场景教学区场景、白天忙时场景商业区场景等; 针对每个场景应用专家规则挑选出

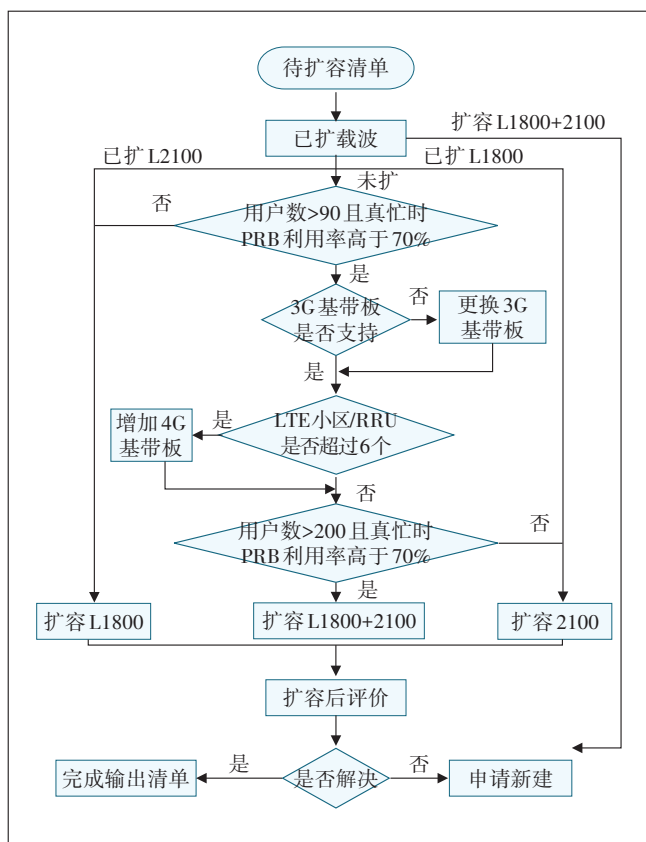


图4 智能扩容决策树

差小区和标杆小区,针对标杆小区计算出场景标杆函数建立流量模型;在质差小区中去除非无线原因导致的质差小区,剩下的无线质差小区结合KK关联规则判断出流量压抑小区,最后算出流量压抑值。

2019年3月评估出流量压抑小区293个,并对这293个小区进行扩容。2019年6月份进行扩容后评估,平均流量增长38.73%。为了验证压抑预测算法的命中率和查全率,对这些样本用传统扩容方法区分哪些是扩容小区作为正样本,哪些是不需扩容小区作为负样本;另外使用流量压抑算法对正负样本进行预测,得到预测算法需扩容小区和预测算法不需扩容小区。针对已扩容272个扇区筛选流量抑制小区校验:命中率=TP/(TP+FP)(本处TP指在传统扩容小区清单中同时也在预测扩容小区清单中,需扩容且扩容后有流量释放的小区,FP指没在传统扩容小区清单中但在预测扩容小区清单中,且扩容后没有流量释放的小区),查全率=TP/(TP+FN)(本处FN指在传统扩容小区清单中,但没在预测扩容小区清单中的小区,且扩容能带来流量释放),命中率基本都在75%以上,精度较高(见表1)。

表1 流量压抑小区的命中率和查全率

| 扩容方案类型 | 检测算法                                      | 总扇区数 | 有效数据记录数(扇区数) | TP  | FP | FN | TN | 命中率/% | 当前查全率精度/% |
|--------|-------------------------------------------|------|--------------|-----|----|----|----|-------|-----------|
| 载波/劈裂  | 激活_忙时流量增益(流量忙时)                           | 272  | 272          | 156 | 89 | 18 | 9  | 63.67 | 89.66     |
|        | 激活_忙时流量增益_(剔除负载不均衡或者无谱效率增益小区)             | 272  | 213          | 149 | 40 | 18 | 6  | 78.84 | 89.22     |
|        | Top10场景_激活_忙时流量增益_(剔除负载不均衡或者无谱效率增益小区)     | 83   | 59           | 39  | 11 | 6  | 3  | 78.00 | 86.67     |
|        | PRB利用率50以下_激活_忙时流量增益_(剔除负载不均衡或者无谱效率增益小区)  | 20   | 18           | 0   | 0  | 15 | 3  | -     | 0.00      |
|        | PRB利用率50~70_激活_忙时流量增益_(剔除负载不均衡或者无谱效率增益小区) | 87   | 73           | 50  | 17 | 3  | 3  | 74.63 | 94.34     |
|        | PRB利用率70以上_激活_忙时流量增益_(剔除负载不均衡或者无谱效率增益小区)  | 165  | 122          | 99  | 23 | 0  | 0  | 81.15 | 100.00    |

## 6 总结

福州联通将持续推进基于用户价值的差异化智能网络容量管理模式创新,通过建立基于AI的流量抑制识别和基于价值的差异化容量管理体系,改变了传统只简单根据小区流量、用户数、PRB利用率作为扩容标准的容量管理模式,精准定位高价值用户,通过保障扩容快速提升了高价值用户感知,达到节约扩容投资成本、提升扩容投资效益的目的。

## 参考文献:

- [1] 工业和信息化部. 促进新一代人工智能产业发展三年行动计划(2018-2020)[EB/OL]. [2020-06-22]. [http://www.cac.gov.cn/2017-12/26/c\\_1122166495.htm](http://www.cac.gov.cn/2017-12/26/c_1122166495.htm).
- [2] 周晓鹏,李程峻. WCDMA同频邻区漏配多配的测试方法分析[C]//信息产业部,2011.

## 作者简介:

陈锋,毕业于福建农林大学,高级工程师,主要从事无线网络优化工作;许绍松,毕业于南京邮电大学,高级工程师,主要从事无线网络优化工作;李张铮,毕业于大连理工大学,工程师,主要从事无线网络优化工作。