

# 基于KNIME workflows 机器学习 预测室内天线外打故障

Indoor Distribution Antenna Outward Emitting Fault Prediction  
Based on Machine Learning of KNIME Workflow

李国博(中国移动通信集团广东有限公司中山分公司,广东 中山 528400)

Li Guobo (China Mobile Communications Group Guangdong Co., Ltd. Zhongshan Branch, Zhongshan 528400, China)

## 摘要:

室内分布系统天线外打故障因难以快速准确定位,对用户体验和网络运维带来了严重挑战。基于KNIME workflows 的机器学习算法预测模型,提出了一种新方法。这种方法能够及时发现并提前解决室内分布系统天线外打故障,实现了从“事后发现”到“事先预测”以及从“全面排查”到“精准定位”的转变,确保室外分布系统天线故障的全面精准定位。

## 关键词:

KNIME workflows; 机器学习; 精准定位

doi:10.12045/j.issn.1007-3043.2024.07.017

文章编号:1007-3043(2024)07-0088-05

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:

Due to the difficulty in quickly and accurately locating indoor distribution antenna outward emitting fault, it poses serious challenges to user experience and network operation and maintenance. Based on machine learning prediction model of KNIME workflow, a new method is proposed, which can timely detect and solve indoor distribution antenna outward emitting fault in advance, achieving a transition from "post discovery" to "pre prediction" and from "comprehensive investigation" to "precise positioning", ensuring the comprehensive and accurate positioning of indoor distribution antenna outward emitting fault.

## Keywords:

KNIME workflow; Machine learning; Precise positioning

引用格式:李国博. 基于KNIME workflows 机器学习预测室内天线外打故障[J]. 邮电设计技术, 2024(7):88-92.

## 0 前言

随着移动通信用户数量的增加和网络技术的进步,室内分布系统天线外打故障已成为无线通信领域的普遍难题。由于天线位置和数量众多,这类故障难以被快速准确地发现和定位,导致网络质量下降和用户体验变差<sup>[1]</sup>。尽管当前通过人工巡检、监控和告警系统能在一定程度上进行故障定位和处理,但存在漏检和误判的问题<sup>[2]</sup>。为解决这一问题,利用机器学习技术,研发了能够提前预测并精准定位天线外打故障

的新方法<sup>[3]</sup>,以及时发现和提前解决故障。

## 1 综合解决方案

机器学习涉及的知识太多,容易使人望而生畏。然而,KNIME workflows 极大地降低了入门的难度,它专注于机器学习思想和应用,无需过多其他专业技能。同时,它是基于Eclipse 开发环境的机器学习工具,采用类似数据流的方式来建立流程,还可以与其他系统(Python、Java、Tableau 等)集成,同时内置各种主流机器学习算法<sup>[4]</sup>。

### 1.1 技术手段

基于KNIME workflows 的机器学习模型,利用人工智

收稿日期:2024-05-19

能核心技术,通过总结经验,发现规律并掌握规则<sup>[5]</sup>。机器学习工作流程如图1所示。



图1 机器学习流程

以监督学习为例,机器学习就是对历史数据进行选定模型训练,不断优化形成可靠模型后,对新的输入数据进行预测<sup>[6]</sup>。机器预测原理如图2所示。

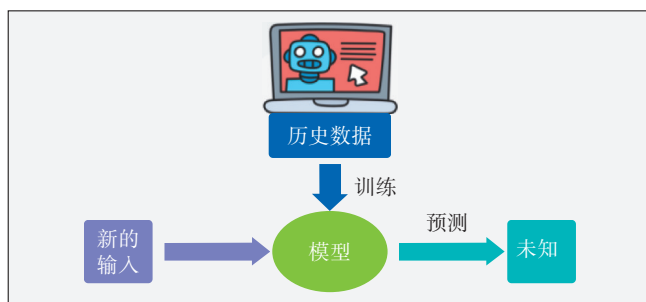


图2 机器预测原理

本次采用基于KNIME工作流机器学习的方式对无线网络室内分布系统天线外打故障情况进行预测<sup>[7]</sup>。KNIME工作流通过采用类似数据流的方式构建流程,实现了模型搭建与数据预测的功能,并将这些流程图形化展示。

a) 数据获取。即从KNIME工作流设置一个Reader节点读取数据,同时可以设置相关数据观察节点进行数据相关性观察。

b) 数据处理。将已有数据划分为训练数据和测试数据,对应训练模型及测试模型<sup>[8]</sup>。

c) 模型训练。选定机器学习模型,设置相关参数。

d) 模型测试。验证及测试模型,将测试数据输入已训练模型。

e) 结果统计及查看预测结果。

f) 模型评价。主要通过ROC曲线图和混淆矩阵判断。

(a) ROC曲线如图3所示。P值(ROC曲线下的面积)越接近1越好。

(b) 混淆矩阵如图4所示。对角线的值越大越好,即准确度(accuracy)越高越好。

## 1.2 技术优点

使用基于KNIME工作流的机器学习模型进行预测,这种方法拥有以下特点与优势。

a) 入门容易,仅需要掌握计算机基础知识和英语,就可以进行机器学习模型的搭建。

b) 通过节点、工作流的方式实现建模,无需学习相关的编程语言。

c) 内置多种主流机器学习算法,方便使用者建模与调测<sup>[9]</sup>。

d) 兼容性强,还可以与其他系统(如Python、Java、Tableau等)集成。

e) 基于机器学习的大数据预测判定具有高可靠性。

本方案主要关注室内分布系统天线外打情况是否存在故障。在方案实施前,主要通过人工分析后台

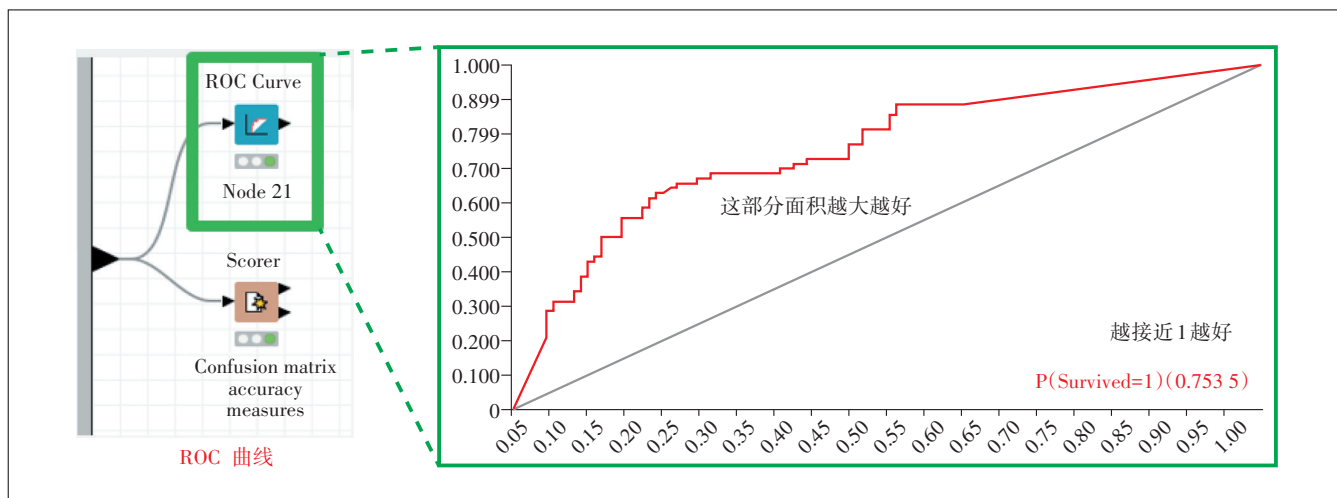


图3 ROC曲线

Survived /...	0	1
0	104	5
1	52	18

Correct classified: 122	Wrong classified: 57
Accuracy: 68.156 %	Error: 31.844 %

混淆矩阵 (confusion matrix)  
对角线的值越大越好

图4 混淆矩阵

网管数据和告警的方式判断故障的存在<sup>[10]</sup>。实施方案后,相对于人工分析判断,故障检测的可靠性得到了提高,同时,在未发生告警的情况下,也能预测并发现问题,实现了从全面排查到精准排查的转变,从而减少了人工成本,同时提高了无线网络排查效率<sup>[11]</sup>。

另一方面,为了验证机器学习的预测准确性和可靠性,有效识别出“真实的”和“虚假的”故障信息,本方案搭建了7种不同算法的预测模型。同时对比不同算法预测的评估效果,选择最适当的模型算法作为室内分布系统天线外打故障预测工具<sup>[12]</sup>,有效降低了误判断的概率,大大提升问题发现率以及故障判断的准确性。

### 1.3 技术情况说明

为了提高机器学习预测的准确性与可靠性<sup>[13]</sup>,本方案搭建了7种不同算法构成的预测模型,分别是:逻辑回归算法、支持向量机算法、决策树算法、随机森林算法、袋装算法、提升算法、贝叶斯算法。

数据特征包括用户面最大激活 UE 数、平均噪声干扰(分贝毫瓦)、上下行总流量(GB)、修正-同时在线用户数、上行 PRB 平均利用率、下行 PRB 平均利用率、最大用户数、RRC 连接建立成功率、ERAB 建立成功率、无线接通率、无线掉线率、ERAB 掉线率、切换成功率、空口上行业务量(GB)、空口下行业务量(GB)、RRC 连接建立最大用户数、无线利用率、上行 RTP 丢包率、下行 RTP 丢包率、低 PHR 小区占比、低 CQI 小区占比、是否故障<sup>[14]</sup>。

模型预测效果评价判断采用 ROC 曲线图结合混淆矩阵的方式实现。

#### 1.3.1 逻辑回归算法

逻辑回归算法的基本原理是尽可能提高所有点分类正确的概率,即尝试将所有点尽可能地远离分类

边界。逻辑回归算法的原理如图 5 所示。

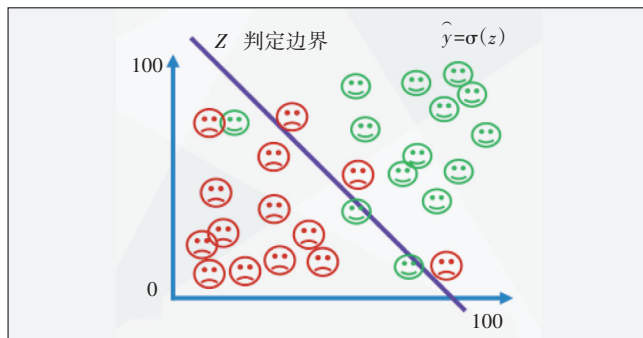


图5 逻辑回归算法原理

#### 1.3.2 支持向量机算法

支持向量机算法的基本原理是尝试最大化由支持向量确定的边界距离,即分类的边界越宽越大越好。支持向量机算法原理如图 6 所示。

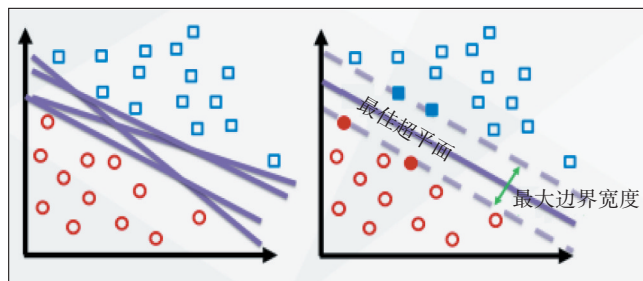


图6 支持向量机算法原理

#### 1.3.3 决策树算法

决策树也叫分类树或回归树,叶子节点给出分类,内部节点代表某个特征,分支代表某个决策规则。构建决策树时通常采用自上而下的方法,在每一步选择一个最好的属性来分裂。“最好”的定义是使子节点中的训练集尽量地“纯”。不同的算法使用不同的指标来定义“最好”。决策树算法原理如图 7 所示。

#### 1.3.4 袋装算法

袋装算法通过从训练集中随机有放回地抽样出  $n$  个样本集,建立  $n$  个决策树模型,然后通过这几个模型的投票来决定结果。通过对几个模型的综合应用,可以在降低偏离的同时降低方差,从而提高预测的准确性和可靠性。袋装算法原理如图 8 所示。

#### 1.3.5 随机森林算法

类似袋装算法,该方法也进行随机有放回抽样,而且对特征也进行抽样,每次抽  $m$  个特征,对特征抽样是为了防止特征之间的相关性对模型的影响。

#### 1.3.6 提升(Boosting)算法



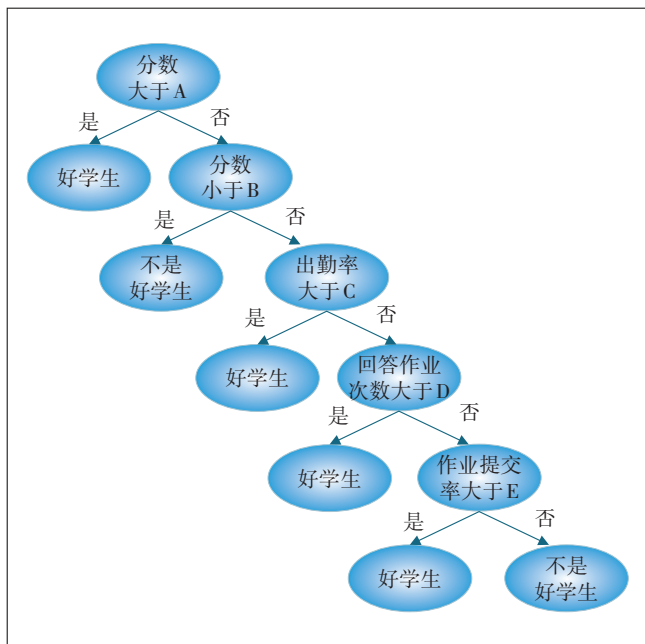


图7 决策树算法原理

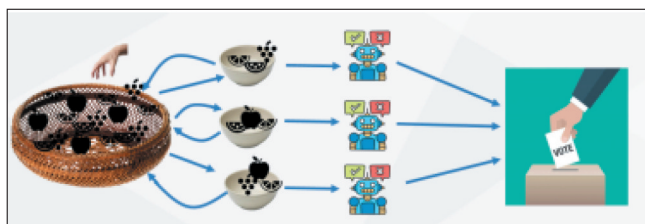


图8 袋装算法原理

与袋装算法类似,提升(Boosting)算法的基本思想、方法是把多个弱分类器集成为一个强分类器。与袋装算法不同的是,袋装算法的每一步都是独立抽样的,而提升算法中每一次迭代则是基于前一次的数据进行修正的,通过提高前一次模型中分错样本在下次抽样中的概率,模型反复学习并优化分类,最终提高模型的准确率。提升(Boosting)算法原理如图9所示。

### 1.3.7 贝叶斯算法

贝叶斯定理描述的是如何使用证据(我们在哪里,在哪个抽样空间中)和先验信念(全集空间,或者

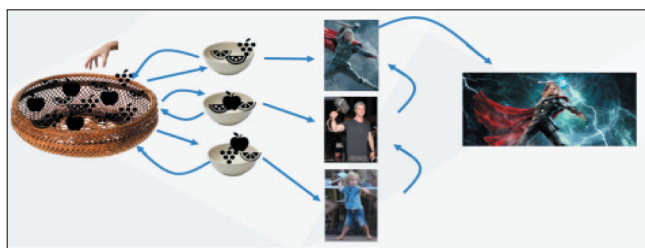


图9 提升(Boosting)算法原理

更大的那个空间)计算事件或假设的条件概率。从先验信念开始,然后获取一些数据并用这些数据来更新信念,更新后的信念被称为后验(信念)。之后获得了更多的数据,旧的后验成又为新的先验,循环重复。在机器学习中,数据就是证据,结果就是信念,目的就是使用数据更新信念。贝叶斯算法原理如图10所示。

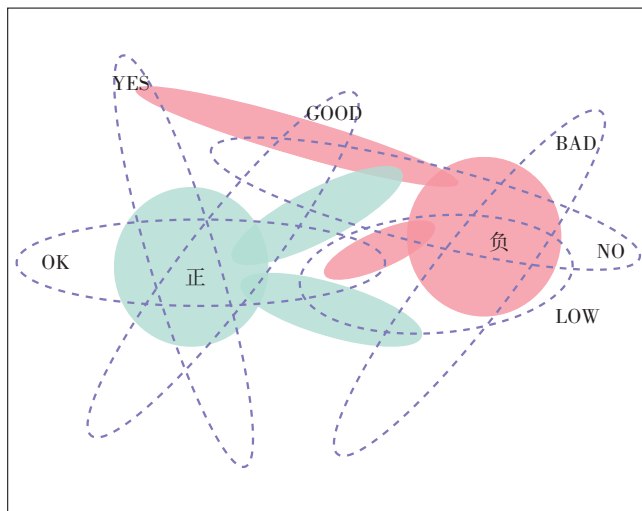


图10 贝叶斯算法原理

## 1.4 模型选择

通过对比7种预测模型发现,在当前数据量和选定特征码的情况下,支持向量机算法的ROC曲线图显示P值等于0.83(最大值是1),而且混淆矩阵准确度(Accuracy)为82%(最大值是100%)。从ROC曲线图P值以及混淆矩阵准确度评估来看,支持向量机算法的表现尚属良好,模型的准确性和可靠性达到实际使用的要求。随着后期不断增加新数据和特征码,并不断进行循环训练学习,该模型还有提升的空间<sup>[15]</sup>。模型选择如图11所示。

## 2 案例应用及分析

某地(市)运营商在预测模型指导下,对20个居民住宅小区的室内分布系统天线外打情况进行现场勘察,最终发现14个住宅小区存在室内分布系统天线外打故障,问题发现准确率达到70%,大大提升了室内分布系统天线外打潜在问题的发现能力及处理效率。此举解决了室内分布系统天线外打故障排查困难、花费工时较长的问题,提高了住宅小区无线网络质量,并降低了无线网络投诉量。部分预测室分系统天线外打故障的住宅小区,整改前后指标对比如表1所示。

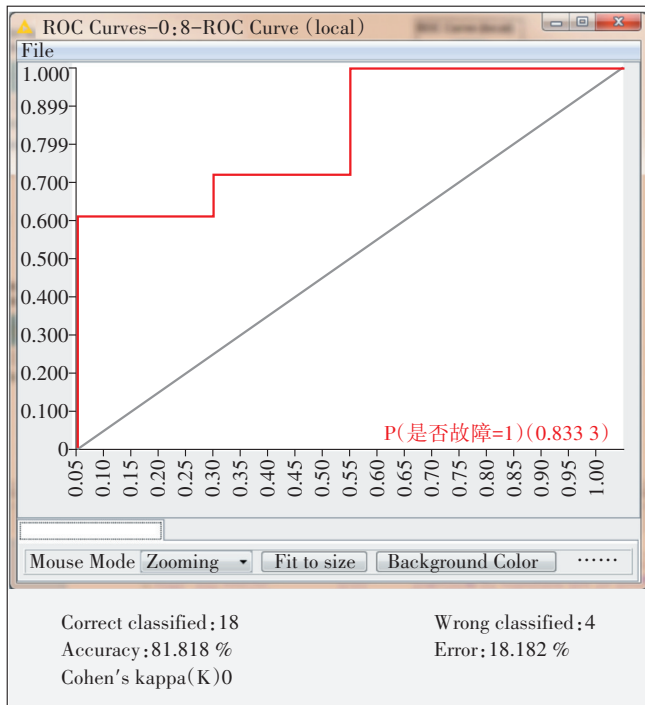


图 11 模型选择

表 1 整改前后指标对比

物业点名称	整改前			整改后		
	日均总上行业务量/GB	日均总下行业务量/GB	单小区平均利用率/%	日均总上行业务量/GB	日均总下行业务量/GB	单小区平均利用率/%
坦洲海伦印象	1.16	11.12	10.8	1.69	18.91	15.9
开发区保利林语二期	0.39	3.42	2.4	1.57	19.88	8.2
开发区凯茵又一城	0.28	3.36	2.5	1.13	15.37	6.9
城区时代云图	0.45	2.41	2.2	3.91	57.72	17.4

从表 1 可以看出,通过基于 KNIME workflows 机器学习模型搭建,对实际住宅居民小区的室内分布系统天线外打情况进行了有效预测<sup>[16]</sup>。在故障处理并恢复后,网络指标得到明显提升。

### 3 总结

KNIME workflows 机器学习降低了机器学习入门的门槛,它专注于机器学习思想和应用。用户无需学习编程语言即可上手。KNIME workflows 还内置多种主流机器学习算法,方便使用者建模与调试,同时其兼容性强,可以与其他系统集成。我们可提供基于 KNIME 的机器学习文档,学习文档以 PPT 的形式提供,帮助使

用者更好地掌握 KNIME workflows 机器学习。此外,该模型还具有使用方便、无需新增硬件设备、无需搭建系统环境、可靠性高等优点,适合推广应用。

### 参考文献:

- [1] 刘昕,苏文俊,曾沂黎. 无源室内分布系统故障探针诊断解决方案[J]. 广东通信技术,2018(4):35-38.
- [2] 高晓勤. 面向 5G 的特殊场景室分网络建设分析研究[J]. 中国新通信,2022(21):23-25.
- [3] 唐浩. 基于智能复合天馈线的室内覆盖系统故障处理维护策略[J]. 通信世界,2018(12):33-34.
- [4] 唐晓芳,周武,吴超. 一种室内分布隐性故障分析算法[J]. 信息通信,2018(3):159-160.
- [5] 石玉伟,邵莹. 基于 RFID 技术的新型室分天线监控系统应用研究[J]. 中国新通信,2019(24):106-108.
- [6] 滕祖伟,周杰华,肖波. 基于 K 最近邻算法的地铁 4G 室分故障定位方法研究[J]. 邮电设计技术,2019(6):52-55.
- [7] 张锐,金奇男. 5G 传统室分故障主动定位新方法的研究[J]. 通信管理与技术,2024(2):41-46.
- [8] 徐东洋. 室内分布系统故障精确定位方案及案例分析[J]. 电信技术,2013(11):85-88.
- [9] 刘贤松,高有利,屠梓浩. 基于多策略融合的室分小区隐性故障发现[J]. 邮电设计技术,2023(4):42-47.
- [10] 张欣旺,曹景阳,徐飞,等. 智慧室分系统研究及应用[J]. 移动通信,2018(9):46-51.
- [11] 杨文,杜彝,陈洁. 基于 AI 的室分故障问题定位方法研究与实践[J]. 电信工程技术与标准化,2020(2):6-10.
- [12] 孙明栋,孙琳,张春莲. 基于时空熵算法和 AI 推理的网络故障根因网元快速界定方法研究[J]. 山东通信技术,2023(1):45-47.
- [13] 田桂宾. 结合 MR 数据设计方案提高 TD-LTE 室分故障排查效率方法的研究[J]. 电信工程技术与标准化,2019(1):16-22.
- [14] 石旭荣,陈辰,刘权,等. 基于机器学习方法的设备监控指标预测[J]. 邮电设计技术,2020(10):15-20.
- [15] 姚克宇,徐俊涛. 通信基站智能化巡检释放人力价值提质增效再加速[J]. 通信世界,2021(6):17-18.
- [16] 葛兆阳,胡勤伟. 高价值室分隐性故障排查研究与实践[J]. 山东通信技术,2023(2):35-37.

### 作者简介:

李国博,高级工程师,学士,主要负责无线网络建设、运维及优化相关专业工作。

