

# AI深度学习在移动网异常小区检测分类中的应用

## Application of AI Deep Learning in Mobile Network Abnormal Cell Detection and Classification

王 勇<sup>1</sup>,滕祖伟<sup>2</sup>,周杰华<sup>2</sup>,肖 波<sup>2</sup>,赵 根<sup>2</sup>(1. 中国联合网络通信集团有限公司,北京 100033;2. 中国联通湖北分公司,湖北 武汉 430020)

Wang Yong<sup>1</sup>, Teng Zuwei<sup>2</sup>, Zhou Jiehua<sup>2</sup>, Xiao Bo<sup>2</sup>, Zhao Gen<sup>2</sup>(1. China United Network Communications Group Co., Ltd., Beijing 100033, China; 2. China Unicom Hubei Branch, Wuhan 430020, China)

### 摘 要:

采用自监督方式检测异常小区、无监督方式聚类问题类型、可视化标注标签的数据挖掘思路,得到带标签的异常小区问题类型数据集。不仅标签数据集可直接应用到网络运营运维领域,而且挖掘问题类型数据集的思路和方法论可广泛推广到其他领域。另外,借助卷积神经网络深度学习算法,将大量的专家经验模型化,构建智能优化引擎,挖掘出人工不易捕捉到的深度信息,更加准确高效地发现网络问题,可扩展性强。

### Abstract:

Using self-supervised method to detect abnormal cells, unsupervised method to cluster problem types and visually labeled data mining ideas, the abnormal cell problem type data sets with labels are obtained. Not only label data set can be directly applied to the field of network operation and maintenance, but also the idea and methodology of mining problem type data set can be widely extended to other fields. In addition, with the help of convolutional neural network deep learning algorithm, a large number of expert experience is modeled, and an intelligent optimization engine is constructed to mine the depth information which is not easily captured by human, so that network problems can be found more accurately and efficiently with strong scalability.

### Keywords:

Deep learning; Abnormal cell; Problem detection; Problem classification

引用格式:王勇,滕祖伟,周杰华,等. AI深度学习在移动网异常小区检测分类中的应用[J]. 邮电设计技术,2019(11):11-15.

## 1 应用背景

当今移动网络复杂多变,特别随着5G和物联网的引入,其复杂性将进一步增加。采用深度学习算法的人工智能解决方案,可以适应复杂的环境,在用户和流量行为不断变化时,将优于任何人工驱动的方案。

移动网异常小区问题类型涉及无线覆盖、干扰、容量等多个方面,传统的网络优化方法是先采集PM和配置等数据,网优工程师再根据专家知识和经验,对各项指标分别设置阈值,来诊断网元是否存在高负

荷、高干扰、覆盖差等异常问题。管理增加了网络复杂性,如何在网络动态变化的情况下,快速发现网络问题,提升运维效率,是目前亟待研究的课题。

基于AI深度学习的移动网异常小区检测分类方案应运而生,它实现了原始数据入库解析自动化,问题分类智能化,诊断结果精准化,管控流程标准化,显著提升日常优化的生产效率和工作效率,起到了降本增效的作用。

## 2 基于AI深度学习的移动网异常小区检测分类

### 2.1 总体方案

基于AI深度学习的移动网异常小区检测分类方

收稿日期:2019-09-03

案由3部分组成,分别是智能建模、深度学习以及智能应用。基于AI深度学习的移动网异常小区检测分类方案如图1所示。

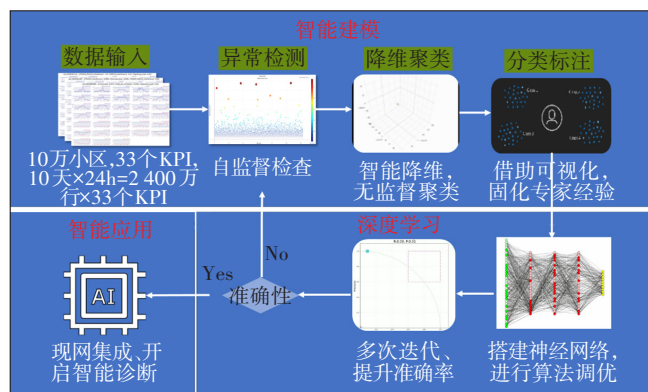


图1 基于AI深度学习的移动网异常小区检测分类方案

a) 智能建模。首先是原始数据收集、数据准备和预处理;然后将预处理好的小区性能数据输入异常检测模块,将输出的结果进行降维聚类处理,并安排有经验的网优工程师对问题进行分类标注,固化专家知识和经验,从而得到带标签的异常小区问题数据集。

b) 深度学习。把带标签的异常小区问题数据集输入搭建好的神经网络深度学习模型进行调优,经过多次迭代和评估,最终得到训练好的神经网络模型及参数。

c) 智能应用。将训练好的神经网络模型部署到实际网络后,每日自动执行移动网异常小区问题类型检测和分类。

## 2.2 智能建模

智能建模流程分为3个主要阶段:异常小区检测、异常小区分类以及专家分类标注。

### 2.2.1 异常小区检测

一般说来,在机器学习领域,不平衡的训练样本会导致训练模型侧重样本数目较多的类别,而“轻视”样本数目较少类别。在异常小区检测场景中,普遍存在的现象是正常样本数量远远大于异常样本数量,因此,当把不平衡的正常样本和异常样本都输入机器学习算法进行训练时,就能把异常样本检测出来。

变分自编码器(VAE—variational autoencoder)是一类在半监督学习和非监督学习中使用的人工神经网络,其架构分为编码器和解码器2部分。其中,编码器神经网络将输入数据转化为隐含空间中更小更紧凑的编码表达,而解码器则将这一编码表达重新恢复为原始输入数据。由于变分自编码器的隐含空间被

设计为连续的分布以便进行随机采样和插值,因此可以有类似正则化防止过拟合的作用。

在样本不平衡的情况下,通过设定不同的重构误差门限,统计VAE所检测出的异常小区数,从而验证检测准确率,如表1所示。

表1 基于VAE的异常小区检测准确率统计

VAE重构误差门限	VAE检测异常小区数	已确认异常小区数	VAE检测准确率/%
7.50E-17	1 590	975	61.3
7.75E-17	1 146	765	66.8
8.00E-17	791	573	72.4
8.50E-17	355	292	82.3
8.75E-17	261	224	85.8
9.00E-17	171	156	91.2
9.25E-17	110	101	91.8
9.50E-17	68	64	94.1
1.00E-16	33	33	100.0

### 2.2.2 异常小区分类

由于移动网小区数据是高维KPI数据,而传统的聚类算法只适用于低维数据。因此,需要将检测出来的高维异常小区数据进行降维,然后再利用传统的已经成熟有效的聚类算法对降维后的低维异常小区数据进行聚类,获得多种不同类型的异常小区。

综合考虑算法有效性和性能,采用LLE降维和K-Means聚类的组合方案。其中,LLE(Locally Linear Embedding)是一种非监督的非线性降维方法,它通过组合数据点的局部属性得到数据的全局结构,而构成这个局部属性的是数据点及它的近邻点的线性组合。其主要思想是利用线性的局部结构来表示全局的非线性结构,通过数据点的局部邻域的相互重叠来近似地表示数据的整体的几何结构。K-Means聚类算法随机选取K个对象作为初始的聚类中心,然后计算每个对象与各个种子聚类中心之间的距离,把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。每分配一个样本,聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。

将采用上述降维聚类算法处理后的异常小区高维数据进行2D降维可视化呈现,异常小区分类效果较好,如图2所示。

### 2.2.3 专家分类标注

将聚类后的异常小区高维KPI数据转换成KPI热

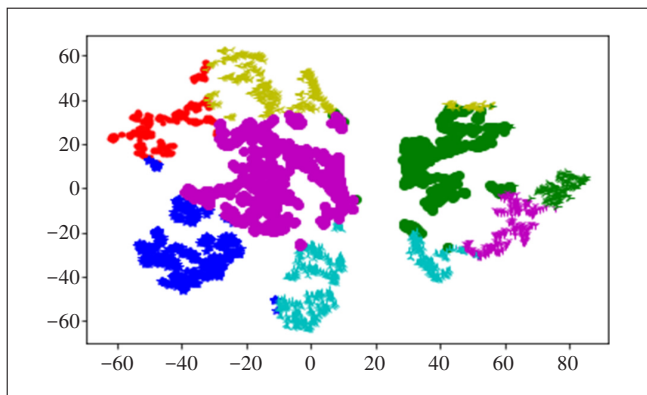


图2 异常小区分类效果可视化

力图,其中,横轴是时间轴,范围是0~23 h,共24个时间点;纵轴是KPI指标。当某个时间点的KPI取值是正常范围时,对应的小方格颜色是绿色,否则对应的小方格颜色是红色。如果KPI取值由正常逐步恶化时,则颜色由绿色逐步变淡,并向红色慢慢演化。KPI

热力图如图3所示。

根据KPI热力图,由有经验的网优工程师手工标注已区分开的异常小区问题类型(例如高负荷、上行干扰等),从而将专家的大量经验进行固化,并最终得到带有标签的移动网异常小区检测数据集。

### 2.3 深度学习

由于卷积神经网络(CNN—Convolutional Neural Network)在准确性和健壮性上的优势,将带有标签的移动网异常小区检测集作为训练数据输入到卷积神经网络进行训练,学习正常小区和异常小区在容量、干扰、覆盖上的特征差异,从而得到预测模型。

具体通过TensorFlow来搭建卷积神经网络实现。首先,设定好卷积神经网络结构(见图4)。然后,将带有标注的异常小区检测数据集划分为训练集和测试集,之后,输入训练集,对卷积神经网络模型进行训练。当模型训练达到一定的准确率后,再次利用测试

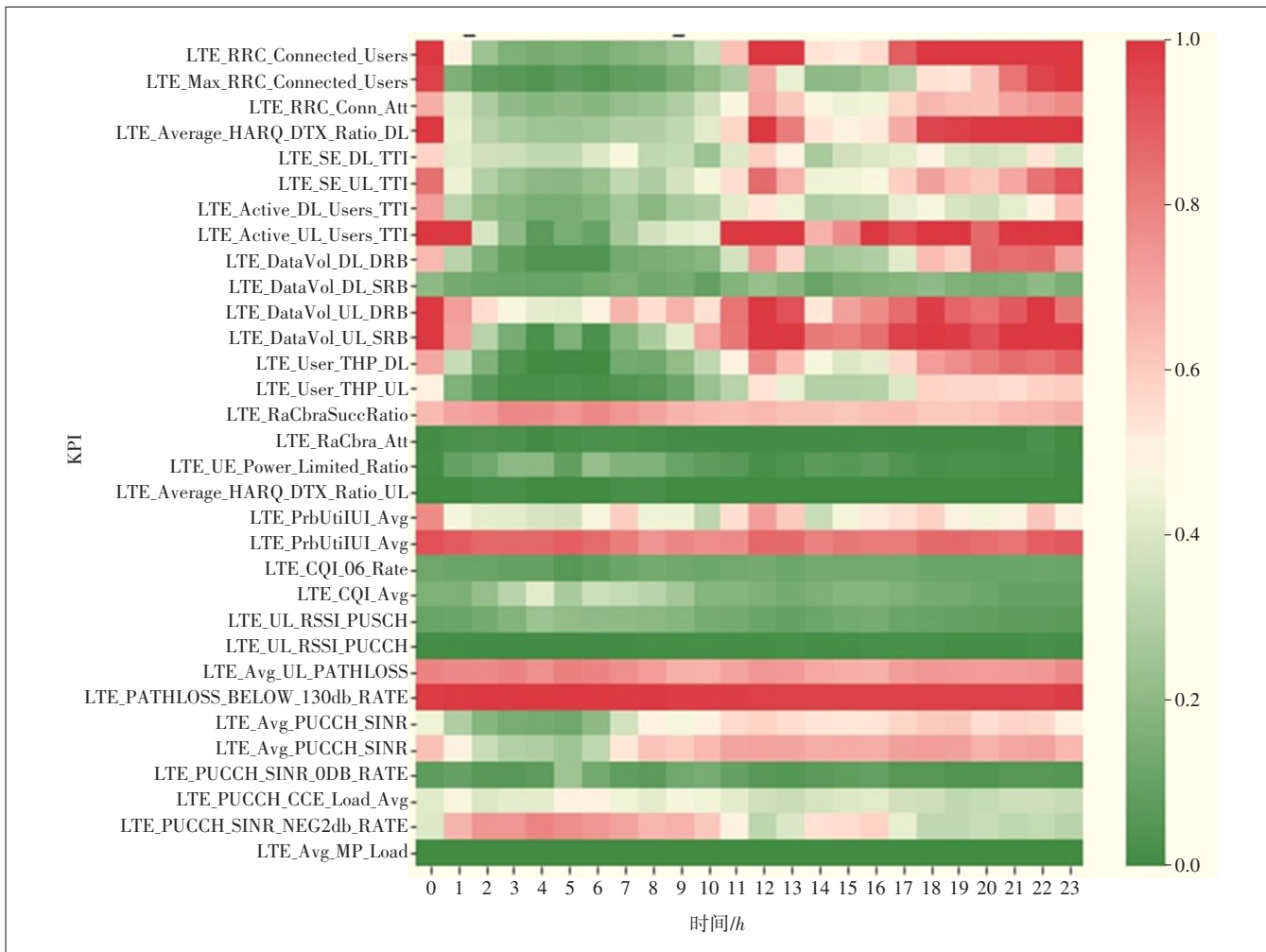


图3 KPI热力图

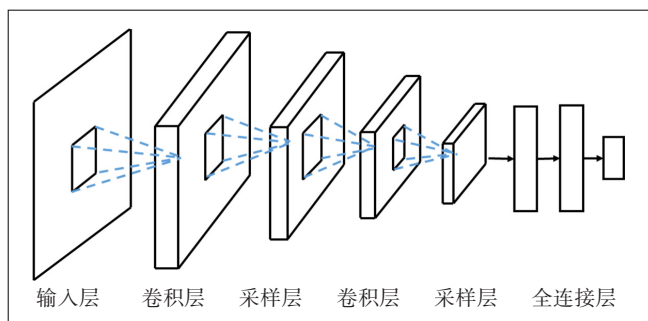


图4 卷积神经网络结构示意图

集对模型进行评估,当模型达到预期标准则输出预测模型。具体的训练和验证阶段的准确度以及损失曲线如图5所示。

## 2.4 智能应用

### 2.4.1 准确性验证

采集武汉联通2019年3月7日4G网络的KPI数据,输入已调测完毕的卷积神经网络模块进行诊断,得到Cell\_load、High\_DL\_utilization、PUCCH\_Interference、DTX\_Coverage等问题类型的诊断概率大于50%的小区,并按各个问题类型数量的20%进行随机采样验证。

针对每类问题类型呈现最相关的2~3类KPI,通过散点图分布特征,判断输出结果是否符合日常优化经验。经验证:异常小区问题检测分类的准确率可达88.37%。异常小区问题检测分类的验证准确率如表2所示。

### 2.4.2 现网部署

采用Docker在Kubernetes上部署深度学习模型。由于深度学习模型的应用组件及所依赖的组件非常多,封装成容器Docker,可实现更轻量级的虚拟化,方

表2 异常小区问题检测分类的验证准确率

异常小区问题类型	小区总个数	验证个数	匹配个数	准确率/%
Cell_Load	207	41	37	90.24
Cell_Load_Interference	9	2	2	100.00
Coverage_DTX	18	4	4	100.00
High_DL_UL_Utilization	14	3	2	66.66
High_DL_Utilization	229	46	36	78.26
PUCCH_Interference	39	8	8	100.00
PUSCH_Interference	12	2	2	100.00
PUSCH_PUCCH_Interference	115	23	23	100.00
Total	629	129	114	88.37

便快速部署,从而显著减少部署的时间成本和人力成本。AI部署组件图如图6所示。

Docker集群至少由3台服务器组成。集群的管理工具为Kubernetes。Docker容器化的应用组件包括:

- Coper\_core:核心容器,封装了Kafka、zookeeper、etcd等基础服务。运行多实例,组成高可用集群。
- Coper\_proxy:把多个coper\_core实例的服务代理成统一接口,供其他服务访问。
- Fetcher:分布式文件下载组件,用以下载性能指标文件,用作数据输入,并提取原始指标存入数据库。
- Timescale DB:时序数据库,存储原始指标。
- Coper\_model:预测模块,用训练模型对数据输入进行处理,输出预测分类结果。
- grafana:集群性能监控。
- Open LDAP:用户管理。

## 3 应用成效

在Docker部署异常小区问题检测分类AI模块的

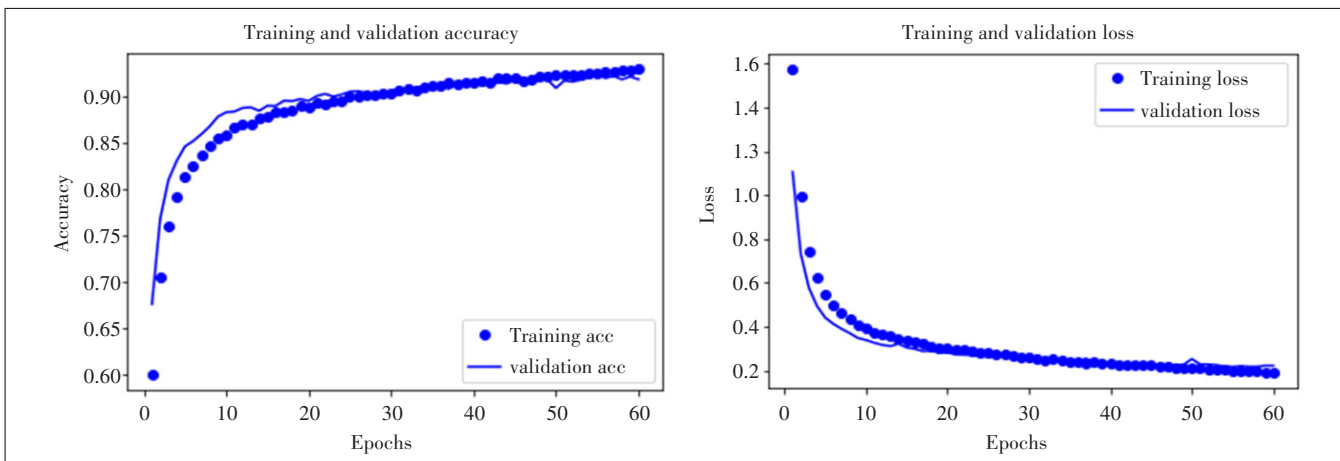


图5 训练和验证阶段的准确度以及损失曲线

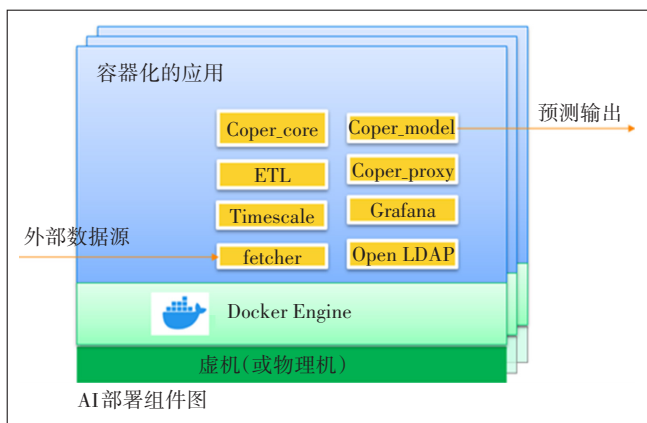


图6 AI部署组件图

基础上,主动对接运维工单系统,自动触发问题工单,串联问题定位、问题处理、问题解决等实施环节,并通

过AI验证优化效果,实现移动网异常小区问题的闭环管控以及网优生产自动化、精细化执行。异常小区问题工单闭环管控如图7所示。

基于AI深度学习的移动网异常小区检测分类方案,引入变分自编码器、卷积神经网络等深度学习算法,集成在中国联通自有大数据平台,并已应用到日常网优生产工作中。通常15 min内能够处理10万个以上的小区,至少节省时间90%以上,全面提升了中国联通网络运维效率。

参考文献:

- [1] 孙志军,薛磊,许阳明.深度学习研究综述[J].计算机应用研究,2012,29(8):2805-2810.
- [2] 张开旭,周昌乐.基于自动编码器的中文词汇特征无监督学习



图7 异常小区问题工单闭环管控

- [J]. 中文信息学报,2013,27(5):1-7.
- [3] 刘堪,袁蕴英.基于自动编码器的短文本特征提取及聚类研究[J].北京大学学报(自然科学版),2015,51(2):282-288
- [4] 展鹏伟,谢小姣.几种降维技术在分类问题中的效果评估[J].科技创新与应用,2018(21):22-23.
- [5] 白俊卿,闫桂荣,王成.利用局部线性嵌入的模式识别[J].西安交通大学学报,2013,47(1):85-89.
- [6] 张少龙,巩知乐,廖海斌.融合LLE和ISOMAP非线性降维方法[J].计算机应用研究,2014,31(1):277-280.
- [7] 左进,陈泽茂.基于改进K均值聚类的异常检测算法[J].计算机科学,2016,43(8):258-261.
- [8] 韩最蛟.基于数据密集性的K均值初始化方法[J].计算机应用与软件,2014,31(2):182-187.
- [9] 骆永健.基于聚类的数据流异常检测算法的研究[D].哈尔滨:哈尔滨工程大学,2010.
- [10] 翟正利,梁振明,周炜,等.变分自编码器模型综述[J].计算机工

- 程与应用,2019,55(3):1-9.
- [11] GOODFELLOW I, BENGIO Y, COURVILLE A.深度学习[M].赵申剑,黎晟君,符天凡,等,译.北京:人民邮电出版社,2017:304-319.
- [12] 贾文娟,张煜东.自编码器理论与方法综述[J].计算机系统应用,2018,27(5):1-9.
- [13] 史加荣,马媛媛.深度学习的研究进展与发展[J].计算机工程与应用,2018,54(10):1-10.

作者简介:

王勇,毕业于北京大学,硕士,现就职于中国联通集团公司运行维护部;滕祖伟,毕业于华中科技大学,硕士,主要从事网络运营等相关工作;周杰华,毕业于解放军重庆邮电学院,学士,主要从事网络运营等相关工作;肖波,毕业于华中科技大学,学士,主要从事网络运营等相关工作;赵根,毕业于中国地质大学,硕士,主要从事网络运营等相关工作。