基于机器学习的网络异常诊断方法

Network Anomaly Diagnostic Methods Based on Machine Learning

李珊珊，陈　凌，刘贤松（中国联合网络通信集团有限公司，北京 100033）
Li Shanshan，Chen ling，Liu Xiansong（China United Network Communications Group Co.，Ltd.，Beijing 100033，China）

摘　要：
主要介绍了基于机器学习的网络异常诊断的模型设计与实现。不同于常见的通过无监督学习进行模型设计及实现的方式，该模型从生产实际出发，结合实际的KPI数据与KPI异常评判的多样性，进行模型设计、训练和优化，实现了网络异常诊断模型的构建。在此基础上，通过集合生产数据在线训练，逐步提升模型精度。

关键词：
KPI；异常诊断；交叉验证；数据闭环
doi：10.12045/j.issn.1007-3043.2025.01.017
文章编号：1007-3043（2025）01-0087-06
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 mainly introduces the model design and implementation of network anomaly diagnosis based on machine learning. Unlike the common way in which model is designed and implemented through unsupervised learning，the model starts from the actual production situation，and combines the actual KPI data and the diversity of KPI anomaly evaluation to design，train and optimize，thus achieving the construction of the network anomaly diagnostic model. On this basis，the model accuracy is gradually improved through online training with production data.

Keywords：
KPI；Anomaly diagnostic；Cross-validation；Data closed loop

引用格式：李珊珊，陈凌，刘贤松. 基于机器学习的网络异常诊断方法[J]. 邮电设计技术，2025（1）：87-92.

0 前言

在通信网络的运维中，网络覆盖类指标、呼叫建立类指标、呼叫保持类指标等各类KPI是日常运维工作的重要数据，这些数据随着时间不断累积，成为一笔重要的数据资产。借助人工智能[1-2]，可以基于这些数据进行网络异常诊断方面的应用，从而预测网络性能是否出现异常，并进行网络问题的排查与调整优化。

收稿日期：2024-11-19

本文将阐述如何从实际情况出发，基于机器学习进行网络异常诊断应用工作，其中包括关键KPI指标筛选、历史KPI数据分析、特征工程处理、结合KPI数据特点进行机器学习模型构建，以及利用实际运行数据进行模型在线优化等内容[3]。

1 背景及问题分析

在获取小区历史KPI数据的过程中，有一些比较关键的KPI。运维专家在长期工作中，会结合这些KPI数据对小区网络进行异常诊断及优化调整等操作，并留有对应的诊断过程和结果记录，相当于给这些KPI

数据做了是否为异常数据的“准标注”。利用这些“准标注”的专家经验数据，结合机器学习方法可以创建一个模型，用于对网络异常的诊断[4-5]。

但是在实际工作中，每个区域，特别是不同应用场景下，KPI指标的波动不同，评估KPI指标是否异常的标准并不是一成不变的。因此，需要结合不同区域、不同场景来进行分析。

从单个KPI指标来看，每个KPI指标都有其自身的数据特点。图1是某个KPI指标的数值分布，可以看到这是个极度偏态的分布，数据都集中在0附近。如果单从分布上看，这大概率是异常情况，但如果从KPI本身的含义上来看，此分布却是合理的，因为在实际生产中这个指标的数值本身就不会太大[6]。

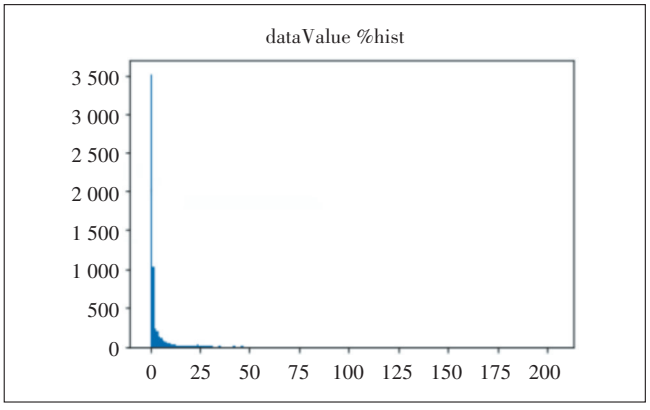

图1　偏态分布的KPI示例

同时，由于网络较少发生异常，在实际生产中“准标注”KPI异常数据的样本量比例偏低。因此，在设计一个KPI异常诊断模型时必须考虑到正负样本失衡的问题，在此基础上先得到一个模型，然后在模型上线后进行更新优化[7]。

2 解决方案

2.1 机器学习整体流程

机器学习模型应用开发和落地一般有2个流程。第1个流程是线下训练，线下训练又包含多个子流程，第1步是业务需求分析，其输出为对应的机器学习任务，这一步是整个课题的重要起点。从业务角度分析，希望能够通过关键KPI数据分析建立起网络异常检测模型，以进行下一步的预案执行，并将这一步作为由数据驱动智慧运维的一个部分。

第2步是根据机器学习任务，从生产数据库或者其他数据源中进行数据提取和清洗等工作。数据是开发机器学习应用的基础，数据质量、数据量、数据可获取性是在机器学习应用开发落地开始时就必须要考虑的3个维度，同时也是直接影响后续机器学习建模、训练效果的重要因素。

第3步是进行机器学习模型训练，这一步包含特征工程、模型选择、模型参数调优、模型交叉测试等任务。模型训练部分的工作比较繁重，具体表现在这部分工作需要通过反复尝试，甚至需要再拉取新的生产数据，以“探索启发式”的方式进行。这部分工作需要通过机器学习平台来完成，而一个智能、面向生产应用的机器学习平台将会带来极大的好处。本研究采用了网络AI平台进行模型研发，网络AI平台从网络业务角度出发，预置集成了面向通信行业场景化的应用模型，这让开发者可以将更多的时间放在业务创新上来，而不是陷入到算法性能调优的重复性工作中。

在第1个流程完成后，通过模型上线部署，机器学习模型开发应用进入到第2个流程，即模型在线服务流程。作为一个面向生产的机器学习模型，必须要能够为上层应用提供服务调用接口，并在实际运行中不断进行在线优化。在机器学习模型在线服务过程中，时间的迁移、数据的变化、应用的调整等因素都可能会导致模型性能发生变化。此外，模型预测精度也会随着时间的推移发生变化，而这种变化往往是负面的，这些都需要通过对模型的监控来及时发现。

当模型性能或预测精度下降到一个门限时，需要触发模型在线更新工作。模型在线更新工作可以通过2种方式进行，其中一种是全量更新的方式，也即通过自动拉取数据，快速走完整个线下训练流程，训练出新的模型，并在A/B测试后进行上线。此方式的特点是模型将得到全面的更新，甚至在线下训练时可以重新设计模型方案，虽然这个过程耗时较长，但其益处是可观的，适合对时间要求不敏感的业务，通过该方式更新的模型上线后将有效拉长模型的线上服务周期。另外一种方式是线上直接进行模型的增量更新，这种方式的特点是更新周期短，并且能够应对突发状况下的快速更新。增量更新可通过较短的时间窗口完成模型的更新，从而快速地将模型从下线“边缘”拉回来。具体采用何种更新方式可结合模型的业务场景进行选择。

2.2 网络异常诊断方案

2.2.1 方案设计

结合机器学习整体流程，以及网络异常诊断研究

的具体特点，进行方案设计。在进行方案设计的过程中，必须充分利用好网络运维专家宝贵的“历史”沉淀经验，将“准标注”数据转换成实际的标签数据，满足有监督训练数据的需求，这样就可以在异常诊断应用启动阶段获得一个比较可靠的“冷”启动模型。

针对异常衡量标准多样化的情况，使用已处理的标记样本直接训练一个针对所有标准的分类模型似乎是一个比较可行的方式。但在实际情况中，每种异常衡量标准的样本数据并不均衡，不同地区和应用场景下的数据特征也不同，因此一个通用的异常检测分类模型在实际业务中并不可行。为此，本文采用对不同区域单独构建模型的方式，并对同一个区域进行场景分离，使同一个场景使用统一的模型，避免不同场景下分类结果有较大差异的情况，建模的整体方案如图2所示。

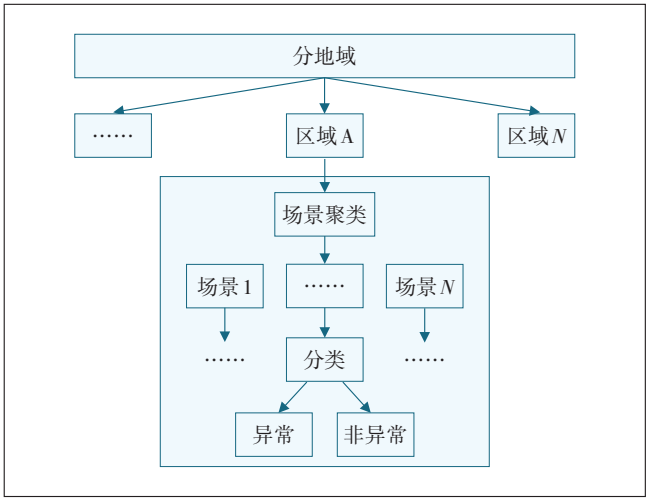

图2　模型整体方案

2.2.2 数据源选择

在基于机器学习的网络异常诊断研究中，数据源的选择至关重要，一般不会将网络中所有的KPI指标都作为特征进行训练（一是基于性能考虑，二是太多无效特征反而会产生噪声）。因此，基于网络中的各类问题，提取一些关键KPI指标作为特征值并参与到机器学习的过程中。本研究提取了无线容量类、质量类、覆盖类、干扰类、切换类的主要指标，并利用Prophet时序算法对KPI指标进行模型训练，得到各指标的动态门限，具体指标及门限如表1所示。

需要说明的是，以上各类KPI指标来自不同的数据源，在具体分析过程中需要进行数据合并，如某些指标不具备采集条件，可根据实际情况进行裁剪。

表1　KPI动态门限

序号	KPI指标	门限（优化后）
1	下行PRB平均占用率/%	39，80
2	PDCCH信道占用率/%	31，80
3	上行PRB平均占用率/%	26，40，80
4	RRC连接重建比例/%	2，50
5	RRC连接建立成功率/%	99.5
6	最大RRC连接用户数/个	175
7	平均RRC连接用户数/个	150
8	UE上下文掉线率/%	0.2，0.9
9	E-RAB掉线率/%	0.15，10
10	用户面下行平均时延/ms	35
11	eNodeB内切换成功率/%	98
12	X2切换成功率/%	95
13	下行PRB双流占比/%	10
14	RSRP大于等于-105的比例/%	74，85
15	平均噪声/dBm	-95
16	平均接入距离/m	500
17	E-RAB建立成功率/%	99
18	CQI≥7占比/%	74，90
19	最大激活用户数/个	100
20	平均激活用户数/个	75

本研究的标签数据为专家分析结果，即人工诊断结果，若专家标签为1则表示异常，0表示非异常。这是一个典型二分类标签，可用于对网络异常的初步定位。实际上，更有用的标签数据是针对问题分类结果的标签数据，如“覆盖类问题—XX原因引起的覆盖类问题—最佳解决方案”等，这在机器学习中可以归结为多分类预测。多分类预测在实际中更有应用价值，但也更加依赖样本数据集的多分类标签。本研究会从二分类起步，最终扩展至网络异常多分类预测。

2.2.3 特征工程

特征工程指把原始数据转变为模型训练数据的过程，其目的是获取更好的训练数据特征，提升机器学习模型的性能。特征工程是机器学习非常重要的环节，一般认为它包括特征构建、特征提取、特征选择3个部分。在本研究中，除了特征值的常规处理外（如异常数据清洗、空值填充、归一化处理等），对于表1所列的各类KPI数据，还进行了特征工程处理。

2.2.3.1 增加指标变化特征

同一小区同一指标的变化往往意味着某些网络异常的发生，这比单纯看某一个指标的绝对值更容易发现问题。因此，本研究增加了指标同比、环比变化值，将其作为新的特征来进行分析。具体而言，对于

每一个指标，增加其环比特征（以3天为一个周期，分别定义今天与昨天、前天、大前天的变化值）以及同比特征（以1周作为一个周期，定义今天与上周同一天的变化值）。这样，对于每个指标，派生出4个新的特征值。

2.2.3.2 增加指标到均值和标准差比值特征

考虑到异常是距离中心比较远的点，所以本研究增加指标到均值和数据分布标准差的比值特征（见图3）。需要进一步说明的是，考虑到同时刻标准差会因为数据少的原因比较敏感，而所有时间段数据会更稳定，因此本文的标准差取自小区所有时段的数据，而且在进行分析时会更关注接近最大值时的异常，不会特别关注在小值处的异常。

2.3 模型的构建

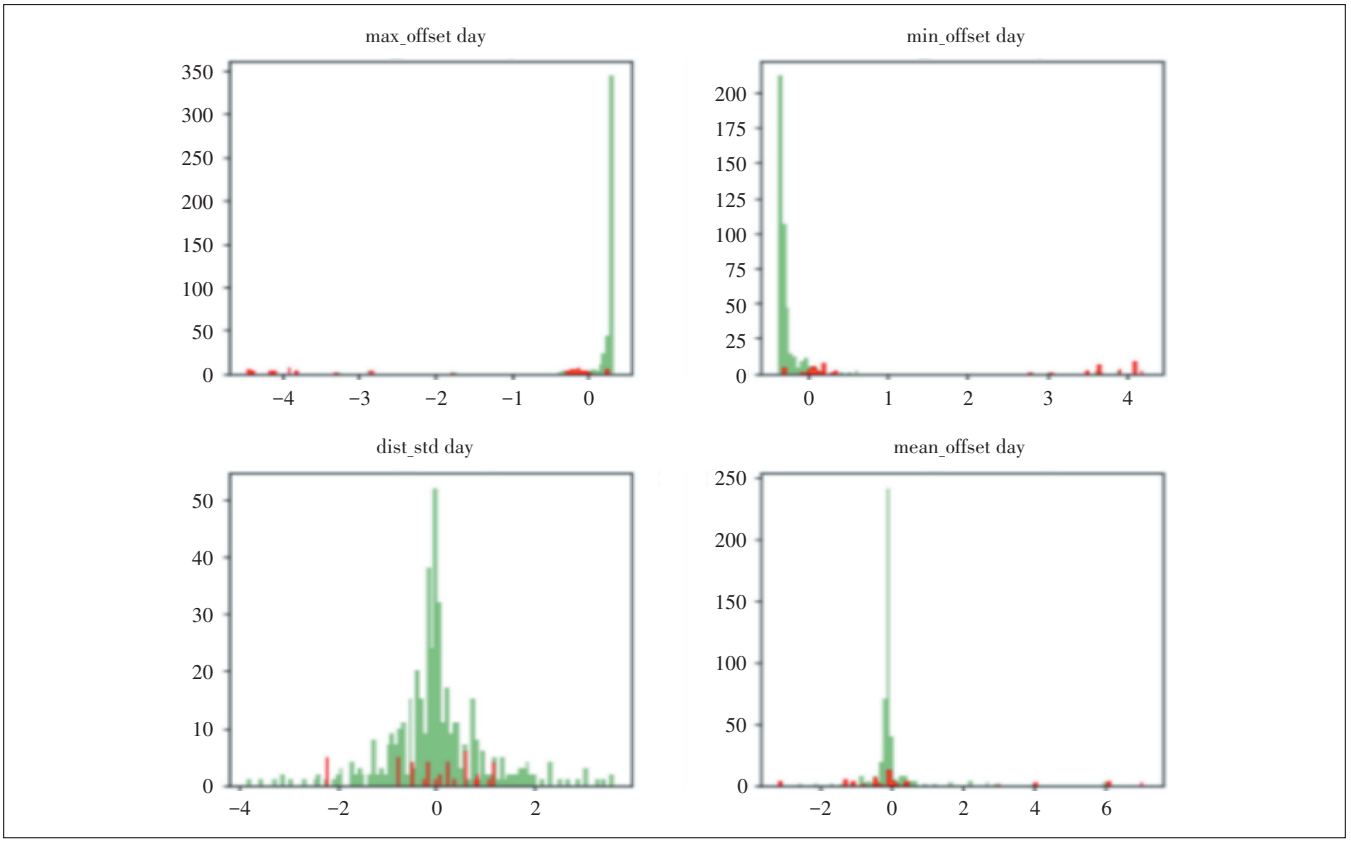

图3　指标到均值和标准差的比值特征分布

2.3.1 模型训练

在设计好模型方案后，需要分步进行模型的构建与实现，其中机器学习算法的选择很关键。

在同一区域数据中，不同应用场景下的KPI波动不同，可以通过结合B域数据以及对KPI指标的聚类来进行应用场景分离，从而找到合适的场景类别数。通过实验对不同的聚类算法进行比较，最后选择K-Means聚类算法进行场景分类，并通过聚类算法的肘分析法来确定场景分类的具体数量，从而得到应用场景分类，如办公楼、商场、学校等（见图4）。

由于抽取的标注样本数据量并不多，不同分类样本的比例严重失调，并且在有监督的机器学习中需要划分训练样本和测试样本，因此在考虑训练数据以及

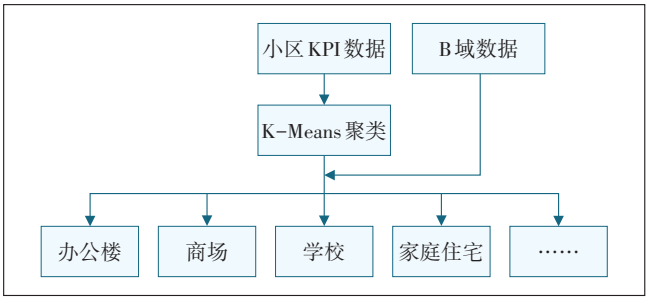

图4　场景分类结果（部分）

模型泛化性能要求的情况下，冷启动阶段不宜使用复杂的模型。

根据以上特点，本研究选择了5种模型，并对数据进行了训练和交叉验证[8]。5种分类模型如表2所示。

表2　分类模型选择

算法	优点	缺点
广义线性类算法（逻辑回归）	简单易懂，速度快，求解容易	无法处理非线性边界问题
决策树（随机森林）	易于理解，可视化，速度快，可混合连续值与离散值，可解决非线性问题	忽略特征之间的相互关联，算法会偏向量较大的分支
贝叶斯（伯努利贝叶斯）	贝叶斯理论更贴近于人类大脑感知事物的工作方式，算法稳定，基本不需要调参	朴素贝叶斯有特征独立的强假设要求，需要预先得知先验概率，分类有一定错误率
K近邻	思想简单，理论成熟，可用于非线性问题，训练时间复杂度低，准确度高，对数据没有假设，对outlier不敏感	计算量大，难处理样本不平衡问题，需要大内存
神经网络（MLP）	可对复杂数据进行建模，分类、预测精度非常高，鲁棒性能好，可解决非线性问题	易过拟合，节点、层数对计算速度影响较大，输出黑盒，无法理解模型，由于解法的随机性，模型稳定性很差

对于上述模型，使用同样的训练数据进行训练并进行交叉验证，将各个模型的性能值进行比较，挑选出适合的机器学习模型[9-10]。各个模型在训练集和测试集上F1值的表现如表3所示。

表3　模型F1值比较

算法	训练集F1	测试集F1
逻辑回归	0.727 206	0.606 249
随机森林	0.814 590	0.775 218
朴素贝叶斯	0.789 954	0.788 483
K近邻	0.734 438	0.644 064
神经网络（MLP）	0.886 697	0.700 630

通过表3中F1值的对比可以看到，模型复杂度较高的神经网络算法在测试集上F1下降很明显，泛化性上有较大损失，模型过拟合现象严重，这也印证了在本文已有标注训练数据和数据量的情况下，不能选择复杂模型作为异常诊断应用的“假设”。同样，随机森林模型训练集效果也很好，但在测试集的表现要差一些，这是因为，在数据量较少的情况下，随机森林也比较容易出现过拟合现象[11]。所以，本文最终选择了朴素贝叶斯模型[12]。

2.3.2　模型的优化

通过模型训练和测试评估，本文选择了朴素贝叶斯模型，并进一步对该模型进行参数优化[13]。相较于复杂的神经网络和树模型算法，朴素贝叶斯模型参数较少，主要有以下几个参数。

a）Alpha。先验平滑因子，默认值为1（表示拉普拉斯平滑）。

b）Binarize。样本特征二值化阈值，默认值是0，None表示模型认为所有特征都已二值化完毕。如果输入具体的值，则模型会把大于该值的部分归为一类，小于的部分归为另一类。

c）fit_prior。是否去学习类的先验概率，默认是True。

d）class_prior。各个类别的先验概率，如果没有指定，则模型会根据数据自动学习，每个类别的先验概率相同，等于类标记总个数（N）的$1/N$[14]。

经过训练，得知Alpha的取值<0.7即可，此处设为0.5；其他参数可直接使用默认值，此时模型的效果已经非常理想，总体测试集F1值的平均表现已经达到了0.793 4。

2.4　模型可靠性分析

由于原始正负样本的比例严重失衡，达到1∶100以上，这种情况下模型肯定会非常偏向于100的部分。因此，本研究进行了样本比例测试以及交叉验证测试，其结果如图5所示。从结果上看，当数据比例在1∶9以内时选择模型的平均F1值最小，但也在0.76以上，证明模型稳定度较好[15]。

2.5　模型上线及更新

在得到“冷启动”模型后，通过网络AI平台进行模

```
for samples in[500,1 000,2 000,4 000]:
    for positive_negative_ratio in [0.01,0.05,0.1,0.2]:
        c1f:BernoulliNB(0.5)
        X,y,ss-prepare_train_data(df_rrc, samples, one_hot_groups=5
        sc=print_cv_result(clf,X,y,result_only =True)
        print("samples={}  positive_weight:{:.2f} Train F1:{:.4f}
```

```
samples=500   positive_weight:0.01  Train F1:0.178 7  Test F1:0.175 6
samples=500   positive_weight:0.05  Train F1:0.606 9  Test F1:0.631 0
samples=500   positive_weight:0.10  Train F1:0.789 7  Test F1:0.793 4
samples=500   positive_weight:0.20  Train F1:0.784 8  Test F1:0.777 9
samples=1 000   positive_weight:0.01  Train F1:0.212 2  Test F1:0.221 0
samples=1 000   positive_weight:0.05  Train F1:0.598 1  Test F1:0.607 8
samples=1 000   positive_weight:0.10  Train F1:0.766 2  Test F1:0.770 6
samples=1 000   positive_weight:0.20  Train F1:0.801 7  Test F1:0.803 4
samples=2 000   positive_weight:0.01  Train F1:0.212 5  Test F1:0.219 3
samples=2 000   positive_weight:0.05  Train F1:0.597 9  Test F1:0.604 5
samples=2 000   positive_weight:0.10  Train F1:0.766 6  Test F1:0.766 9
samples=2 000   positive_weight:0.20  Train F1:0.861 9  Test F1:0.859 9
samples=4 000   positive_weight:0.01  Train F1:0.219 8  Test F1:0.223 0
samples=4 000   positive_weight:0.05  Train F1:0.550 3  Test F1:0.552 4
samples=4 000   positive_weight:0.10  Train F1:0.771 4  Test F1:0.771 5
samples=4 000   positive_weight:0.20  Train F1:0.850 2  Test F1:0.850 1
```

图5　选择模型的交叉验证F1值

型的部署上线，并和生产系统对接实际的生产数据，对小区性能是否异常进行预测。预测的结果同时又可以通过运维专家的反馈，进行校验。长期运行后，更多高质量的数据将得到沉淀。

通过定时任务，可周期性地进行模型的在线训练、优化与更新，达到数据闭环和模型闭环的效果，其整体流程如图6所示。

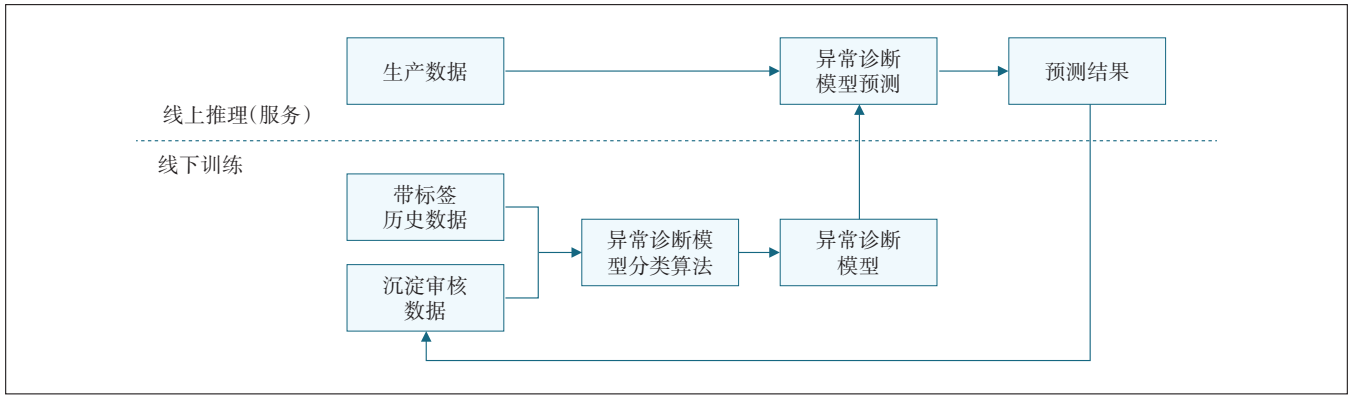

图6　模型的线上服务数据闭环

3 结语

本文基于机器学习进行KPI异常诊断的应用研究，不同于常见的通过无监督学习进行异常诊断的方式。在进行方案设计的过程中，本文充分分析了KPI数据特点，并从实际工作角度考虑了不同区域、不同应用场景下多变的异常评判标准，再结合运维专家宝贵的“历史经验”沉淀下来的“准标准”数据，设计和构建了“冷启动”KPI异常诊断模型。模型上线后，通过更多实际数据的沉淀继续进行模型的在线优化及更新，不断提高模型的运行效率。

参考文献：

[1] 卢奇，科佩克. 人工智能[M]. 林赐，译. 2版. 北京：人民邮电出版社，2018.

[2] 古德费洛，本吉奥，库维尔. 深度学习[M]. 赵申剑，黎成君，符天凡，等，译. 北京：人民邮电出版社，2017.

[3] DU M，LI F F，ZHENG G N，et al. DeepLog：anomaly detection and diag-nosis from system logs through deep learning[C]//ACM Conference on Computer and Communications Security (CCS). Singapore：ACM，2017：1-14.

[4] 陈俊. 基于大数据机器学习技术的IT运营分析系统建设[J]. 计算机时代，2018(3)：85-88.

[5] 刘凡平. 神经网络与深度学习应用实战[M]. 北京：电子工业出版社，2018.

[6] 盛骤，试式千，潘承毅. 概率论与数理统计[M]. 4版. 北京：高等教育出版社，2018.

[7] LECUN Y，BENGIO Y，HINTON G. Deep learning[J]. Nature，2015，521(7553)：436-444.

[8] AGGARWAL C C. Outlier analysis[M]. 2nd ed. Cham：Springer，2016.

[9] ALMAGUER-ANGELES F，MURPHY J，MURPHY L，et al. Choosing machine learning algorithms for anomaly detection in smart building IoT scenarios[C]//2019 IEEE 5th World Forum on Internet of Things (WF-IoT). Limerick：IEEE，2019：491-495.

[10] AMARBAYASGALAN T，JARGALSAIKHAN B，RYU K H. Unsupervised novelty detection using deep autoencoders with density based clustering[J]. Applied Sciences，2018，8(9)：1468.

[11] 王满，谢亚楠. 随机森林算法在运营商告警推送中的应用研究[J]. 工业控制计算机，2017，30(6)：108-109.

[12] ANITESCU C，ATROSHCHENKO E，ALAJLAN N，et al. Artificial neural network methods for the solution of second order boundary value problems[J]. Computers，Materials & Continua，2019，59(1)：345-359.

[13] ASGHAR M Z，KHAN A，BIBI A，et al. Sentence-level emotion detection framework using rule-based classification[J]. Cognitive Computation，2017，9(6)：868-894.

[14] ALPAYDIN E. Combined 5×2cv F test for comparing supervised classification learning algorithms[J]. Neural Computation，1999，11(8)：1885-1892.

[15] BENGIO Y，GRANDVALET Y. No unbiased estimator of the variance of k-fold cross-validation[J]. Journal of Machine Learning Research，2004(5)：1089-1105.

作者简介：
李珊珊，毕业于厦门大学，工程师，硕士，主要从事网络AI开发和算法研究工作；陈凌，毕业于复旦大学，硕士，主要负责算法技术研究工作；刘贤松，毕业于武汉大学，硕士，主要从事网络AI能力研发管理工作。