



# 基于TXF-Net联合架构模型的基站故障预测算法研究

徐焕然, 王磊, 汪浩博, 代稷坤, 杜涵

(中国移动通信集团设计院有限公司业务与IT开发部, 北京 100089)

**摘要:** 基站作为移动通信网络的关键接入节点, 其故障会直接导致网络服务中断与用户体验下降。传统方法难以有效利用多源异构特征(高维时序告警与静态统计特征)并捕捉其复杂映射关系。为此, 提出了一种基于特征融合的级联式深度学习模型TXF-Net(Transformer-XGBoost fusion network)。该模型首先利用Transformer编码器对基站的告警时间序列进行深层编码, 捕捉其长程依赖与动态模式, 并输出高层次的时序语义特征向量。随后, 将该向量统计特征共同输入至XGBoost模型进行最终决策。实验表明, TXF-Net模型在Top1和Top3准确率等关键指标上均显著优于单一模型, 验证了该方法在精准判定基站故障根因方面的有效性。

**关键词:** 基站故障预测; Transformer; 模型级联; 特征融合; 时间序列分析

**中图分类号:** TN929.5

**文献标志码:** A

**doi:** 10.11959/j.issn.1000-0801.2025301

## 0 引言

随着5G网络的高速发展和深度部署, 截至2025年3月月末, 我国基站数量已达到439.5万<sup>[1]</sup>, 基站设备的复杂化与运行环境的多样化导致故障发生率显著上升。传统依赖人工经验的故障处理模式修复时间较长, 大量用户通信中断引发投诉率上升。告警数据作为基站异常状态的直接反馈, 包含故障发生前后的关键信息, 但现有技术未充分挖掘告警数据的时序规律与语义关联, 也缺乏不同模型在故障预测场景下的适配性分析, 导致预测精度与泛化能力不足。

当前基站故障相关研究主要分为两类: 一类是从基站建设状态、频率干扰、激活状态等业务维度进行人工分析<sup>[2]</sup>, 此类方法需要收集大量基站相关数据, 且依赖人工分析; 另一类是构建根

因分析算法模型, 例如, 以XGBoost等为代表的树模型结合一些特征工程对故障分析进行建模<sup>[3]</sup>, 树模型擅长处理结构化特征和捕捉特征间的非线性关系, 但对时序数据中的长距离依赖和动态模式建模能力有限; 而基于深度学习的时序建模方法, 如采用LSTM捕捉告警序列规律, 但未融合文本特征与统计特征, 在复杂故障场景下鲁棒性不足。此外, 现有研究缺乏多模型横向对比分析, 也未探索模型融合联合架构, 难以充分发挥不同模型的优势。

本文研究价值体现在3个方面。(1) 模型对比层面: 通过XGBoost、FM、LSTM、Transformer等多种模型的对比实验, 明确不同模型在基站故障预测场景下的适配性, 为运维实践提供模型选择依据。(2) 技术创新层面: 提出TXF-Net模型框架, 融合Transformer的时间序列建模能力和XGBoost的统计特征/结构化特征建模能

力,解决单一模型难以兼顾时序规律与统计特征的问题。(3)运维实践层面:基于真实数据验证联合模型的有效性,预测准确率提升至67.3%,Top3准确率达到90.1%,可为运维人员提供精准的故障预测,定位基站故障及问题。

## 1 相关技术与模型基础

### 1.1 传统机器学习模型

#### (1) 极端梯度提升树 (XGBoost)

XGBoost<sup>[4]</sup>基于梯度提升决策树框架,通过集成多个弱分类器提升预测性能,支持对缺失值的自动处理与正则化优化,能有效捕捉特征间的非线性关联。在基站故障预测中,XGBoost可利用告警频次、时间特征等统计信息,适合处理结构化特征输入,但难以捕捉告警序列的时序依赖关系。

#### (2) 因子分解机 (FM)

FM<sup>[5]</sup>通过因子分解技术建模特征间的低阶交互关系,解决传统线性模型无法处理特征交叉的问题。对于基站告警数据中的“告警类型-时间窗口”“告警频次-故障类型”等交叉特征,FM能有效挖掘其隐含关联,但其对时间序列的建模能力较弱,无法捕捉告警发生的先后顺序规律。

### 1.2 深度学习时序模型

#### (1) 长短期记忆 (LSTM) 网络

LSTM网络<sup>[6]</sup>通过门控机制(输入门、遗忘门、输出门)解决传统循环神经网络(recurrent neural network, RNN)的梯度消失问题,可捕捉时间序列的长期依赖关系。在基站告警序列处理中,LSTM网络能挖掘故障前告警的时序演变规律,如“故障前30 min射频单元告警频次上升”,但仅能单向捕捉时序信息,对双向时序关联的建模能力不足。

#### (2) Transformer

Transformer<sup>[7]</sup>基于自注意力(self-attention)机制,可并行处理时序数据并计算不同时间步特

征的关联权重,能捕捉非连续告警事件的隐性关联(如“故障前1 h网元连接中断+故障前10 min小区闭塞告警”)。相较于LSTM网络,Transformer对长序列的建模能力更强,但缺乏对低阶特征交互的捕捉能力。

### 1.3 联合建模思路

模型融合是解决异构数据建模的有效手段,通过将不同来源、不同类型的特征有机结合,可显著提升模型性能。本文采用的特征融合策略属于“早期融合”与“中期融合”的结合。

(1) 利用Transformer将时序数据转换为固定维度的特征向量。

(2) 将时序特征向量与原始结构化特征进行拼接。

(3) 利用XGBoost对融合后的特征进行最终决策。

这种策略既保留了原始特征的信息,又融入了时序数据的高层语义表示,实现了两种模型的优势互补。

## 2 融合架构设计及特征工程

### 2.1 整体架构

TXF-Net架构分为3个核心阶段:时序特征提取阶段、特征融合阶段和最终决策阶段。TXF-Net融合架构如图1所示。

(1) 时序特征提取阶段:利用Transformer模型对原始时序数据进行编码,生成包含时序动态模式的特征向量。

(2) 特征融合阶段:将Transformer生成的时序特征与原始统计特征、业务特征进行拼接,形成完整特征集。

(3) 最终决策阶段:利用XGBoost模型对融合特征进行训练,实现故障根因的分类预测。

### 2.2 Transformer时序特征提取模块设计

#### 2.2.1 输入序列构建与特征工程

为充分捕捉基站故障发生前后告警数据的时

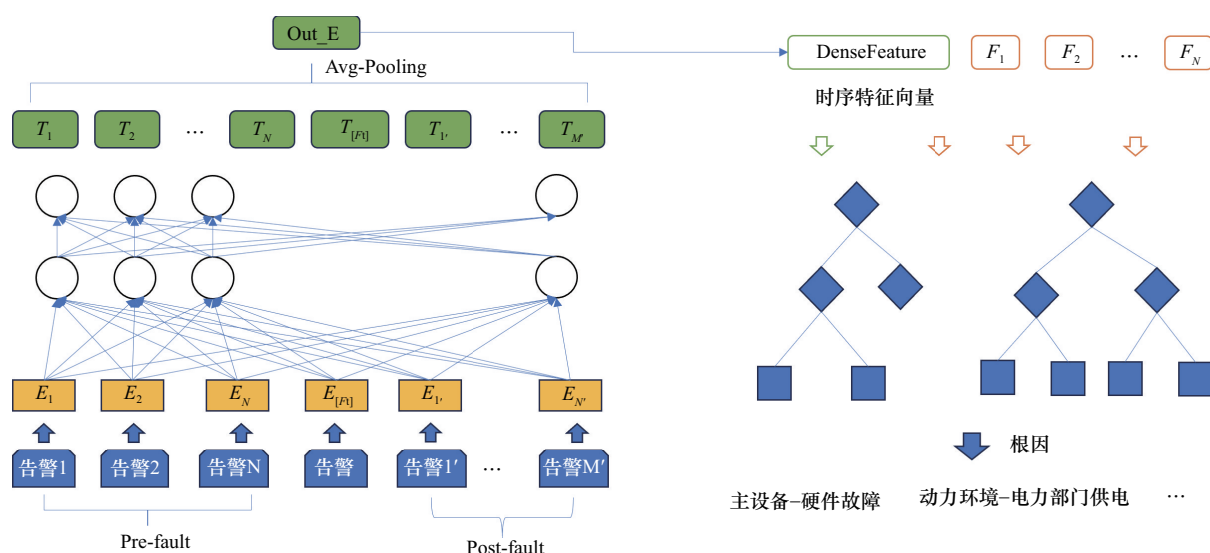


图1 TXT-Net融合架构

序关联与动态模式，本文研究以“故障事件”为核心锚点，构建包含故障前后告警数据的时间序列，并通过多维度特征工程将非结构化时间信息与告警语义信息转化为模型可识别的向量表示。

### 2.2.1.1 时间序列构造逻辑

以每条故障记录为目标样本，按基站身份ID与时间维度关联历史告警数据，构建“故障前T1告警序列-故障时刻-故障后T2告警序列”的完整时序窗口。对窗口内的告警数据按时间戳升序排序，形成长度可变的告警序列；若序列长度超过预设阈值，截取固定长度告警数据；若序列长度不足，则通过零向量填充至固定长度，确保输入序列维度一致性。

### 2.2.1.2 多维度特征编码方案

针对告警数据中的“告警类别”与“时间信息”两类核心信息，设计分层特征编码策略，具体包括以下3个方面。

(1) 告警类别语义编码：基站告警数据包含“射频单元交流掉电告警”“小区不可用告警”等20余类告警类型，采用标签编码（label encoding）将离散告警类别映射为整数索引，再通过嵌入层（embedding layer）转换为稠密向量，捕捉告警类别间的语义关联，如“射频单元交流掉电

告警”与“射频单元维护链路异常告警”的向量距离被约束为相近值。

(2) 时间特征结构化编码：时间信息是故障预测的关键前兆，本文研究通过分段离散化与周期性编码结合的方式，将时间特征转化为高区分度向量，具体处理逻辑如下。

① 时间段划分：基于基站故障发生的时间分布规律，将一天24 h按[0,6,12,18,24]划分为“凌晨”“早上”“下午”“晚上”4个时间段，采用独热编码转换为4维向量，再通过嵌入层压缩为8维特征向量，以捕捉不同时段故障高发的规律（如“凌晨”时段市电波动导致的动力环境故障占比更高）。

② 星期划分：将每星期按星期一到星期日划分为7个周期段，同样通过“独热编码+嵌入层”转换为8维特征向量，适配基站设备维护周期。

③ 周期性正弦/余弦编码：为保留时间的连续周期性特征，对“小时”与“星期几”进行正弦-余弦编码，计算式如下：

$$\text{hour\_sin} = \sin\left(2\pi \times \frac{\text{hour}}{24}\right) \quad (1)$$

$$\text{hour\_cos} = \cos \left( 2\pi \times \frac{\text{hour}}{24} \right) \quad (2)$$

$$\text{weekday\_sin} = \sin \left( 2\pi \times \frac{\text{weekday}}{7} \right) \quad (3)$$

$$\text{weekday\_cos} = \cos \left( 2\pi \times \frac{\text{weekday}}{7} \right) \quad (4)$$

该编码方式可有效捕捉“每天同一时段”“每周同一天”的故障周期性规律，避免传统独热编码导致的维度稀疏问题。

(3) 特征融合与序列构建：对单条告警数据，将“告警类别嵌入向量+时间段嵌入向量+星期嵌入向量+小时正余弦编码+星期正余弦编码”进行拼接；再按时间顺序排列所有告警的特征向量，最终构建维度为 $[\text{batch\_size}, \text{seq\_len}, \text{d\_model}]$ 的输入时间序列。

### 2.2.2 Transformer 编码器结构设计

为聚焦时序特征提取能力，本文研究采用纯编码器架构的Transformer模型（无解码器模块），通过多层自注意力机制捕捉告警序列中的长距离依赖关系，具体结构与参数配置如下。

(1) 位置编码层：由于Transformer模型不具备天然的时序感知能力，在输入序列进入编码器前，需添加位置编码以保留时序信息。本文研究采用可学习的位置嵌入（learned positional embedding），维度与后续自注意力层的模型维度（ $\text{d\_model}$ ）保持一致；将位置嵌入与告警特征向量拼接后，通过线性层映射后得到，确保时序信息与语义特征在同一维度空间可交互。

(2) average pooling层：编码器最后一层输出维度为 $[\text{batch\_size}, \text{seq\_len}, \text{d\_model}]$ 的时序特征张量，其中每个时间步（token）对应一条告警的高层语义特征。为得到固定维度的时序特征向量，本文研究采用全局平均池化（average pooling）策略，对所有时间步的特征向量取均值，计算式如下：

$$\text{pooled\_feat} = \frac{1}{\text{seq\_len}} \sum_{t=1}^{\text{seq\_len}} \mathbf{x}_t \quad (5)$$

其中， $\mathbf{x}_t$ 为第 $t$ 个时间步的特征向量，最终得到维度为 $[\text{batch\_size}, \text{d\_model}]$ 的全局时序特征向量，该向量融合了整个告警序列的动态模式与周期性规律。

(3) 多分类预测头（预训练任务）为使Transformer模型学习到与故障预测相关的时序特征，采用有监督预训练策略：在average pooling层后添加一层FFN，以交叉熵损失函数（cross-entropy loss）训练模型预测故障类别。该预训练任务可引导Transformer聚焦于对故障预测有价值的时序模式（如“射频单元告警连续出现+凌晨时段”与“动力环境故障”的强关联），为后续特征提取提供高质量的时序表示。

## 2.3 XGBoost特征融合与决策模块设计

为充分整合Transformer提取的高层时序特征与基站故障预测相关的细粒度统计特征，本文研究设计“双源特征融合+XGBoost决策”的模块架构，通过多维度特征互补提升故障根因分类精度，具体设计如下。

### 2.3.1 Transformer时序特征

基于第2.2节训练完成的Transformer编码器，对所有训练集与测试集样本的告警序列进行特征提取：将样本的告警序列输入最优预训练Transformer模型，截取average pooling层输出的全局时序特征向量（即 $[\text{batch\_size}, \text{d\_model}]$ 维度张量），直接作为XGBoost的高层时序特征。该特征向量已融合告警序列的长距离依赖关系与周期性规律，是传统统计特征无法覆盖的动态信息。

### 2.3.2 细粒度统计特征

针对基站故障根因与告警发生频率、关键告警时序位置、时间属性的强关联特性，设计4类细粒度统计特征，通过业务经验与相关性分析筛选出高区分度特征，具体包括以下3点。

(1) 多时间窗口告警频次特征基于故障发生时刻，划分多个关键时间窗口（故障前1天、3天、



1周等),每个窗口内统计各类告警的出现次数,形成“窗口数×告警类别数维”特征。

(2) 关键告警位置特征通过 Pearson 相关性分析即卡方检验等方式,发现,故障前最后一个告警与故障后第一个告警的类型故障根因的相关性极高,是强预测因子。对此设计告警类型编码:将故障前最后一个告警、故障后第一个告警的类型分别通过标签编码转换为 One-Hot 向量(对应告警类别)。

(3) 故障时间属性特征基于故障发生的时间信息,构建结构化特征,包括以下3点。

① 小时特征:故障发生小时(0~23)的数值特征(One-Hot 编码)。

② 星期特征:故障发生星期几(0~6)的数值特征(One-Hot 编码)。

③ 时段特征:故障发生时段(“凌晨”“早上”“下午”“晚上”)的 One-Hot 编码。

### 2.3.3 特征拼接

将两类特征进行横向拼接,形成融合特征向量:

融合特征=[时序特征向量]+[统计/业务特征向量]

(6)

假设时序特征维度为  $d_{\text{model}}$ ,统计/业务特征维度为  $n$ ,则融合后的特征总维度为  $d_{\text{model}}+n$ 。

## 3 实验方案设计

### 3.1 数据来源与格式

实验数据来源于某运营商2025年1月—3月的运维数据,包含两类核心数据集。

(1) 故障数据集:约6万条记录,字段包括工单ID、基站ID、故障发生时间、故障类型。

(2) 告警数据集:约300万条记录,字段包括基站ID、告警发生时间、告警标题。

数据集按照训练集:验证集:测试集8:1:1的方式划分。

### 3.2 评价指标

采用以下指标评估模型性能。

(1) 准确率(accuracy):正确分类样本占总样本的比例。

(2) Top3 准确率:预测概率前3的类别包含真实类别的比例。

### 3.3 对比实验设计

为了更充分地对比模型选型对结果的影响,同时选取了FM模型与LSTM模型进行对比实验,共设置3组对比实验。

(1) 对比非序列模型,选取两个经典模型,树模型XGBoost和因子分解机FM,另外在特征工程上也会对比一下基础特征工程,以及增加了若干特征后的效果。

(2) 对比序列模型,同样选取两个经典模型,LSTM和Transformer,另外在特征工程上也会对比一下基础特征工程,以及增加了若干特征后的效果。

(3) 对比最优非序列模型和最优序列模型,以及联合建模后的效果。

### 3.4 各模型设置

(1) XGBoost 模型

超参数设置:树深度=5,学习率=0.1,迭代次数=500,子样本比例=0.8。

(2) FM 模型

超参数设置:因子维度=8,学习率=0.01,迭代次数=100,正则化系数=0.001。

(3) LSTM 模型

网络结构:2层LSTM(隐藏层维度64),Dropout=0.2。

训练策略:Adam优化器,学习率= $3 \times 10^{-4}$ ,批大小=32,早停策略(验证集5轮无提升则停止)。

(4) Transformer 模型

网络结构:2层编码器(注意力头数=4,隐藏层维度64),Dropout=0.2。

训练策略：Adam 优化器，学习率= $3\times 10^{-4}$ ，批大小=32，早停策略（验证集 5 轮无提升则停止）。

### 3.5 实验环境

硬件环境：72 核 Intel(R) Xeon(R) Gold 6240 CPU @ 2.60 GHz，单核 NVIDIA Tesla V100 GPU (32 GB)。

软件环境：Python 3.7，PyTorch 1.13.1，Scikit-learn 1.0.2。

## 4 实验结果与分析

对 3 组实验分别进行数据结果的呈现和分析。

基础模型性能对比见表 1，对非序列化的经典模型进行了对比，在基础特征集上（仅包含一个月内若干窗口不同类型告警数量的统计值），实验结果显示基于 XGBoost 的树模型结构在性能上显著优于 FM，这表明树结构在当前故障类别极度不均衡条件下，有明显的优势。XGBoost 优化版本在 Base 的基础上，进行了更多的特征工程，加上了更长的时间窗口和时间、星期等信息，准确率较 LSTM 提升 4.5 个百分点，体现了特征工程在该场景下的重要性。

表 1 基础模型性能对比

| 模型                | Top1 准确率 | Top3 准确率 |
|-------------------|----------|----------|
| FM-Base           | 50.4%    | 71.0%    |
| XGBoost-Base      | 61.2%    | 83.2%    |
| XGBoost-Optimized | 65.7%    | 88.5%    |

序列模型性能对比见表 2，对经典的序列模型进行了对比评测。从实验数据可知，LSTM-Base 模型表现最弱，受限于单向时序建模能力，即使加长告警时间窗口（故障前报警数从 10 增加到 20），也没有明显效果；Transformer-Base 凭借自注意力机制，在加长时间窗口以及进行更多的特征引入之后（包括增加告警后特征）性能提升明显，充分验证了 Transformer 架构及相关特征优化对基站故障时序特征的捕捉优势。

表 2 序列模型性能对比

| 模型                    | Top1 准确率 | Top3 准确率 |
|-----------------------|----------|----------|
| LSTM-Base             | 56.7%    | 80.3%    |
| LSTM-LW               | 56.8%    | 80.5%    |
| Transformer-Base      | 56.8%    | 80.4%    |
| Transformer-LW        | 58.5%    | 82.5%    |
| Transformer-Optimized | 61.4%    | 84.1%    |

最后，多模型性能对比见表 3，最终 Top1 准确率达 67.3%，较单一 Transformer 提升 5.9 个百分点，较 XGBoost 提升 1.6 个百分点，说明联合架构能同时利用时序编码与特征交互的优势，弥补单一模型的不足；其 Top3 准确率超过了 90%，在实际生产中可以为运维人员提供良好的数据支撑以及自动化能力。

表 3 多模型性能对比

| 模型                    | Top1 准确率 | Top3 准确率 |
|-----------------------|----------|----------|
| XGBoost-Optimized     | 65.7%    | 88.5%    |
| Transformer-Optimized | 61.4%    | 84.1%    |
| TXF-Net 联合建模          | 67.3%    | 90.1%    |

## 5 结束语

本文提出的 TXF-Net 架构通过特征融合策略，有效结合了 Transformer 的时序建模能力和 XGBoost 的结构化特征处理优势，在设备故障根因分析任务中取得了优异性能。实验结果表明：（1）特征融合策略能够充分发挥两种模型的优势，显著提升分类准确率；（2）Transformer 提取的高层时序特征与传统统计特征、业务特征具有互补性；（3）良好的特征工程对根因定位模型的效果起到了决定性的作用。

未来研究可从以下方向展开：（1）探索更精细的特征融合策略，如注意力机制加权融合；（2）引入动态特征选择机制，自适应调整各类特征的权重；（3）尝试端到端的联合学习模型，进行更深入的领域建模；（4）将该融合架构扩展到其他工业场景，如预测性维护、质量检测等领域。



## 参考文献:

- [1] 工信部. 2025年第一季度通信业经济运行情况[R]. 北京: 工业和信息化部, 2025.
- [2] 张涛, 尹劲松, 翟耀威, 等. 一种5G零业务小区根因分析模型[J]. 通讯世界, 2022, 29(5): 166-168.
- [3] 贾子寒, 王西点, 徐晶, 等. 基于机器学习的基站故障根因分析方法研究[J]. 电信工程技术与标准化, 2023, 36(9): 1-8.
- [4] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 785-794.
- [5] Rendle S. Factorization machines[C]//Proceedings of the 2010 IEEE International Conference on Data Mining. Piscataway: IEEE Press, 2011: 995-1000.
- [6] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [7] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.