

# 基于自注意力 Transformer 编码器的 多阶段电力系统暂态稳定评估方法

房佳姝, 刘崇茹, 苏晨博, 林晗星, 郑乐\*

(新能源电力系统国家重点实验室(华北电力大学), 北京市 昌平区 102206)

## Multi-stage Transient Stability Assessment of Power System Based on Self-attention Transformer Encoder

FANG Jiashu, LIU Chongru, SU Chenbo, LIN Hanxing, ZHENG Le\*

(State Key Lab. for Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University),  
Changping District, Beijing 102206, China)

**ABSTRACT:** The artificial intelligence method has made considerable achievements in power system transient stability assessment (TSA). A conventional deep network is generally regarded as a "black box" model, which limits the dependability of intelligent algorithms for practical engineering applications. Furthermore, conventional methods have limited ability to capture the dynamic evolution process of the power system. To solve the above problems, this paper introduces a multi-stage transient stability assessment method based on Transformer encoder, and missing alarms can be effectively reduced with multi-stage predictions. Compared with conventional methods, the Transformer model presents good interpretability, whose attention structure visualizes the internal work of the deep neural networks, so the model can adaptively identify and focus on the important features. Meanwhile, this paper utilizes multi-moment information to construct feature spaces. Compared with other networks, the Transformer gains a global receptive field via attention mechanism, and thus the state change of power system can be characterized quickly and accurately. Simulation on IEEE-39 bus system shows that the proposed method presents good interpretability compared to common data-driven models, showing higher accuracy of transient stability assessment and stronger robustness under data pollution.

**KEY WORDS:** power system; transient stability assessment (TSA); Transformer model; self-attention; attention visualization

**摘要:** 人工智能方法在电力系统暂态稳定评估研究中已经取

得了一定的成果。常规深层网络普遍被视为“黑盒”模型,这限制了智能算法在实际工程应用中的可信赖性;同时,常规算法对电力系统时序信息的提取能力不足。针对以上问题,构建基于 Transformer 编码器的多阶段暂态稳定评估方法,其多阶段预测能够有效降低失稳漏判率。和常规算法相比,Transformer 模型具有良好的可解释性,其注意力机制引导模型自适应识别并聚焦于关键特征,在一定程度上揭示深层网络内部工作决策过程。此外,采用多时刻信息构建特征空间,Transformer 通过注意力机制实现全局感受野,使模型快速捕获电力系统前后时刻间的状态依赖。IEEE-39 节点系统上的仿真结果表明,所提方法相比常见数据驱动模型具有更高的暂稳评估准确性,呈现出良好的可解释性,并在数据污染时依然维持较高的性能。

**关键词:** 电力系统; 暂态稳定评估 (TSA); Transformer 模型; 自注意力; 注意力可视化

## 0 引言

随着新能源在电力系统中占比不断提高,电力电子设备大量增多,电力系统的“双高”特性愈发凸显,由于计算节点的急剧增加以及电力电子设备与控制的紧密耦合,传统的暂态稳定评估方法,如时域仿真法<sup>[1]</sup>、直接法<sup>[2]</sup>等已经无法满足电网调度对快速性和准确性的需求。时域仿真法具有良好的模型适应性和结果可靠性,但计算效率低。直接法计算速度快,但结果偏保守。

人工智能技术的快速发展推动了基于数据驱动的暂稳评估方法。广域测量系统(wide area measurement system, WAMS)的成熟和应用,实现了电力系统动态信息的实时监控<sup>[3]</sup>,为数据驱动的

基金项目: 中央高校基本科研业务费专项资金项目(2021MS005)。

Project Supported by Fundamental Research Funds for the Central University (2021MS005).

暂态稳定在线评估提供了数据保障。

基于数据驱动的暂态稳定评估方法具有强大的特征提取和数据挖掘能力,且能快速准确地在线生成分析结果,目前已取得了很大的研究进展。当前主要采取2类机器学习策略,一类采用“人工特征降维+浅层网络”<sup>[4-7]</sup>,另一类采用能够实现输入和输出信息“端对端”映射的深层网络<sup>[8-17]</sup>。

采用浅层网络<sup>[4-7]</sup>进行暂稳评估的研究,如文献[4]筛选出边界样本作为决策树(decision tree, DT)输入样本从而减少运算时间,文献[5]采取多组支持向量机(support vector machine, SVM)的输出作为下层SVM的融合特征,文献[7]组合多棵基于随机向量的决策树组成随机森林(random forest, RF)。以上文献均需要前置特征工程对输入信息降维处理。然而暂稳分析领域的特征工程依赖于人工经验,缺乏统一标准,其特征选择方式会很大程度上影响模型性能<sup>[18]</sup>。

深层网络不再需要前置特征降维,打破了浅层网络发展的瓶颈<sup>[9-10]</sup>。目前已有研究使用深度置信网络(deep belief network, DBN)<sup>[10-11]</sup>、堆栈自动编码器(stacked autoencoder, SAE)<sup>[12-14]</sup>等深层模型预测系统暂稳状态。文献[10-12]利用大量无标注样本无监督训练模型降低仿真时间。文献[11]将数据映射入二元表达空间以降低评估难度。文献[13]引入代价敏感损失函数提高SAE对失稳样本的评估能力,文献[14]在SAE中嵌入卷积结构提高评估性能。然而,以上深层网络中不具有记忆或标志数据序列关系的神经元结构,因此无法提取特征间的序列关系<sup>[19]</sup>。

目前,深度学习主要采用卷积神经网络(convolution neural network, CNN)<sup>[15-16]</sup>和循环神经网络(recurrent neural network, RNN)<sup>[17]</sup>提取电网动态信息间的时序关系。文献[16]基于1D-CNN提取暂态时序特征,文献[17]采用特殊结构的RNN网络平衡评估精度和响应时间。以上方法均具有较高的泛化能力和准确率,然而捕捉特征间的序列依赖关系时,CNN的“感受野”层叠结构加剧了运算复杂度,RNN串行输入会引发高运算成本和梯度问题。此外,以上序列深层网络可解释性较差,无法展示模型是如何基于特征空间和模型结构输出对应的决策。

2017年文献[20]首次提出以注意力运算为核心的Transformer模型,给处理序列特征提供了全新

的深层网络架构。当前Transformer已成功应用于自然语言处理<sup>[21]</sup>、语音识别<sup>[22]</sup>和计算机视觉<sup>[23]</sup>等领域,在机器翻译领域,基于Transformer的模型已在多个公开数据集上取得目前最好的翻译效果<sup>[24]</sup>。

Transformer模型的核心为多头自注意力机制。注意力机制(attention mechanism)给高价值信息分配高权重,本质上是对信息处理资源的高效分配<sup>[20]</sup>。其自适应注意力权重分布体现了输出数据与输入序列特征间的相关性。常见深层算法中的“黑箱”特性限制了智能算法在实际工程中的可信赖性,使电网调度部门对智能算法给出的决策抱有一定顾虑。目前,可解释性往往由树状结构算法引入<sup>[25]</sup>,决策树基于局部贪婪建立模型,不会对模型进行回溯检测,使模型收敛于局部最优而非全局最优,生成的特征重要度不稳定。而Transformer的可解释性来自于算法结构中的注意力运算,在模型迭代降低损失的过程中生成较为稳定的特征注意力分布,展示特征对于决策的贡献度,并给后续的数据收集方向提供建议。

相比采用故障清除时刻的量测信息进行单次预测<sup>[10-17]</sup>,多阶段预测能够有效降低失稳漏判率,由于系统的失稳特征会随着时间推移逐步明显,使失稳趋势更易被模型识别。在多阶段预测暂稳状态时,文献[26-27]在故障清除时刻后采用滑动窗逐次预测暂稳状态;文献[28]将故障前10周波数据作为原始输入特征量的稳态运行部分,在故障持续阶段多次暂稳评估,并在故障清除时刻确认最终的暂稳评估结果。本文的特征空间包含故障发生前、故障发生时刻和故障清除后的3个时刻断面信息,因注意力机制能够提取特征空间内序列前后状态之间的关系,本文实验证明采用多时刻信息作为特征空间能够有效提高模型准确率,有利于模型对电力系统动态响应的表达。

本文提出一种基于Transformer编码器的多阶段暂态稳定评估方法,模型采用发电机电气信息对系统暂稳状态进行多阶段评估以减少失稳漏判,因失稳状态下的机组失稳特性会随着时间逐渐明显,使其更易被模型识别。此外,相比其他深层算法,Transformer模型的全局感受野实现了特征间的直接交互,其注意力分布体现了特征的价值度,通过注意力机制对关键特征的自适应聚焦,给深层模型带来了良好的可解释性。在IEEE-39节点系统上的仿真结果表明,本文所提方法在模型准确度、可解

释性、以及面对污染数据时的鲁棒性 3 个方面相比常规数据驱动模型具有更优越的性能。

本文所提的基于 Transformer 编码器的多阶段暂态稳定评估方法主要有如下特点。

1) 建立面向电力系统暂态稳定评估问题的 Transformer 模型, 修改标准 Transformer 的编码器—解码器结构, 构建多层编码器与 Softmax 分类层组成的 Transformer 分类器, 模型嵌入层(Embedding Layer)表征时间断面下的电力信息, 实现了不同时序间电气信息的直接交互, 输出暂稳评估结果。

2) 为提高暂稳评估的准确性, 降低失稳漏判, 提出多阶段暂态稳定评估方法。采用稳态状态, 故障发生时刻, 和故障清除后的当前时刻信息构建特征空间。由于注意力机制会提取特征空间内的序列前后状态之间的关系, 因此本文将稳态和故障发生时刻信息加入特征子集, 给模型自注意力运算提供了更丰富的上下文信息, 提升模型性能。

3) 本文重点分析了 Transformer 模型中的注意力机制, 通过解构不同输入特征维度和不同故障场景下的注意力分布, 直观展示深层网络对重要电气

特征和机组的注意力聚焦过程, 验证特征间注意力分布与电力系统暂态状态的相关性, 给数据驱动方法下的暂态稳定评估带来了一定可解释性。改进了现有的基于深层网络的暂稳评估模型中准确率较高但可解释性不足的问题。

4) 考虑到真实电力系统场景中存在数据污染, 通过展示注意力分布对高价值特征的聚焦, 表明算法有效抑制了异常信息对暂稳预测的干扰。本文所提方法不仅具备更高的准确性, 在数据污染下依然呈现出较好的鲁棒性, 这说明本文算法适用于实际电力系统。

## 1 Transformer 注意力网络

本文构建 Transformer 编码器作为分类器核心, Transformer 编码器由  $N$  个相同的子模块(transformer block)堆叠组成, 如图 1 所示, 子模块包含多头注意力层(multi-head attention)和前馈神经网络(feed-forward network)2 个主要部分, 引入残差连接(residual connection)和层归一化(layer normalization, LN)以防止梯度退化并加速算法收敛。

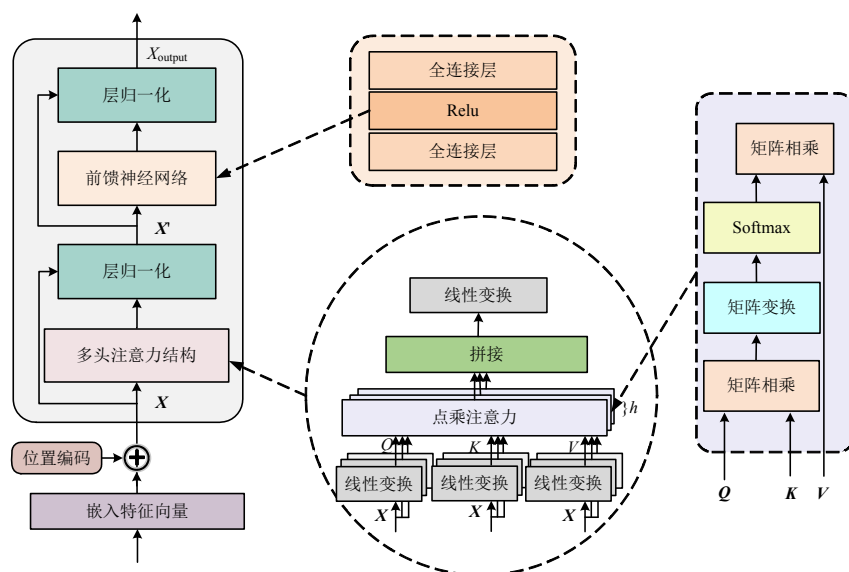


图 1 Transformer 编码器子模块基本架构

Fig. 1 Basic structure of Transformer encoder block

### 1.1 位置编码

位置编码(position embedding)用来描述特征间的相对位置关系, 与嵌入特征层(embedding layer)叠加。Transformer 模型中位置编码计算公式如下:

$$\begin{cases} P_E(p_{os}, 2i) = \sin(p_{os}/10000^{2i/d_{model}}) \\ P_E(p_{os}, 2i+1) = \cos(p_{os}/10000^{2i/d_{model}}) \end{cases} \quad (1)$$

式中:  $p_{os}$  代表序列长度;  $i$  代表特征向量的维度下

标;  $d_{model}$  为特征长度。因为正弦和余弦函数具有周期性,  $P_E(p_{os}+k, n)$  可以表示为  $P_E(p_{os}, n)$  的线性变化, 从而识别特征间的相对位置关系。

在分析电力系统动态变化中的时序数据时, 位置编码用于标记特征在时间维度上的先后顺序。

### 1.2 多头自注意力运算层

自注意力运算模块为 Transformer 编码器的核心单元。多头注意力(multi-head attention)机制可以

视为注意力集成。如图1中所示,通过并行 $h$ 次注意力计算,得到多个参数独立的“表示子空间”,扩展了模型专注于不同特征的能力。

输入矩阵 $X \in R^{s \times d_k}$ 经过3种不同的线性变换得到为查询矩阵 $Q$ ,键矩阵 $K$ 和值矩阵 $V$ 。

$$\begin{cases} Q = W^Q X \\ K = W^K X \\ V = W^V X \end{cases} \quad (2)$$

式中: $W^Q, W^K, W^V \in R^{d_{\text{model}} \times d_k}$ ;  $W^Q, W^K, W^V$ 将嵌入向量 $X$ 从 $d_{\text{model}}$ 维度映射到 $d_k$ 维空间。

多头自注意力计算公式如下:

$$\begin{cases} A_{\text{attention}}(Q, K, V) = F_{\text{softmax}}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \\ h_{\text{ead}i} = A_{\text{attention}}(Q_i, K_i, V_i) \\ M_{\text{ultiHead}}(Q, K, V) = C_{\text{oncat}}(h_{\text{ead}1}, h_{\text{ead}2}, \dots, h_{\text{ead}i})W^O \end{cases} \quad (3)$$

点乘注意力为特征融合运算。 $\text{Softmax}$ 函数 $F_{\text{softmax}}(\cdot)$ 有利于反向传播梯度计算,同时将结果平滑至0-1区间。采 $\sqrt{d_k}$ 用尺度化以避免点积过大造成 $F_{\text{softmax}}(\cdot)$ 梯度过低。多组 $d_k$ 维空间内的注意力分布采用 $C_{\text{oncat}}(\cdot)$ 拼接,经过权重矩阵 $W^O$ 变换得到最终的注意力层输出。

注意力权重公式为 $A_{ij} = F_{\text{softmax}}(Q_i K_j^T / \sqrt{d_k})$ ,  $A_{ij}$ 表示不同特征间的注意力权重,在提取机组注意力分布时,将机组的全部量测信息整合作为该机组的特征以进行注意力运算,并由此得到注意力分布图。因此,数据经过自注意力运算层后的输出视为特征图中的所有特征的加权和。相比卷积和循环网络,自注意力运算实现了不同位置特征间的直接点乘融合,不受特征间距离影响,增强特征表示。从这个角度来说,自注意力机制具备全局感受野,从而可以在数据污染下维持更多信息<sup>[29-30]</sup>,而卷积和循环结构则是在空间或时间上处理局部域,需要通过重复卷积和逐步传播信号的方式捕获特征间的远程依赖关系,进而增加了算法面向污染数据时的优化难度<sup>[31]</sup>。此外,也有文献[3]指出自注意力机制对于污染数据的鲁棒性来自于其动态调整感受野(receptive field)以调整对不同信息的注意力倾向。因此,文章将自注意力机制引入TSA,使其在面向受污染的电气信息时依旧能维持较好的鲁棒性。

### 1.3 前馈神经网络

前馈神经网络层由2层全连接网络构成,每层

对输入向量线性映射,中间的隐藏层使用 ReLu 函数激活,前馈神经网络公式如下:

$$F_{\text{FN}}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (4)$$

式中: $x$ 为注意力层归一化后的输出向量; $W$ 为权重向量; $b$ 为偏置项。

### 1.4 层归一化运算与残差连接

如图1(a)所示,特征矩阵经过多头注意力层和前馈神经网络后均会采用层归一化与残差连接。

层归一化运算中每层神经元共享均值和方差,使层输入归一为标准正态分布,增强反向传播信息流动性。其公式如下:

$$\begin{cases} \mu = \frac{1}{T} \sum_{i=1}^T X_i^2 \\ \sigma = \sqrt{\frac{1}{T} \sum_{i=1}^T (X_i - \mu)^2} \\ L_N(x) = \alpha \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \end{cases} \quad (5)$$

式中: $T$ 为该层神经元的个数; $\mu$ 为均值; $\sigma$ 为标准差; $X_i$ 表示特征向量 $X$ 的第 $i$ 维; $\alpha, \beta$ 为可学习的模型参数,通过反向传递更新; $\varepsilon$ 为防止除数为0设置的小数。

残差连接结构使梯度传递更为高效,避免模型退化。经残差连接的Transformer Block内部数据如下式:

$$X' = L_N(M_{\text{MultiHead}}(X)) + X \quad (6)$$

$$X_{\text{output}} = L_N(F_{\text{FN}}(X')) + X' \quad (7)$$

式中: $X_{\text{output}}$ 和 $X$ 分别为Transformer Block的输出和输入矩阵; $X'$ 为前馈网络层的输入矩阵; $L_N$ 为层归一化运算。

## 2 基于Transformer的多阶段暂态稳定评估

### 2.1 Transformer 评估结构设计

为构建面向暂态稳定评估问题的Transformer分类模型,需要修改标准Transformer的编码器-解码器<sup>[20]</sup>结构。电力系统评估更适合划分为分类问题,因此建立多层编码器与Softmax函数分类层组成的Transformer分类器。

Transformer被广泛应用于自然语言处理领域,其嵌入特征(Embedding)层采用skip-gram索引将单词映射为稠密连续的向量,通过查表方式统一原始词信息的特征维度,这与本文工作有2个主要区别。

1) 词向量是单通道(channel)向量,而电力系统

具备功角、电压、有功、无功等特征, 为多通道向量, 多通道向量无法在保留特征序列信息的情况下直接通过 Transformer 进行注意力融合。

2) 经过 skip-gram 的单词会自动获得统一的词向量维度, 而电力系统中不同节点(如发电机节点和母线节点)可能具有不同的特征维度。

对此, 我们在 Embedding 层采用  $1 \times 1$  卷积对原始高维电气信息进行数据融合, 这种操作解决了上述问题。

1) 对原始高维电气量直接进行注意力运算会破坏特征图中的序列信息, 且模型无法聚焦于某一类具体的电气特征。而  $1 \times 1$  的卷积核将 4 通道数据(功角、电压、有功、无功)降维为单通道数据, 不仅减少 Transformer 中注意力交互的计算量, 更是保留了数据序列信息, 且使模型能够通过注意力运算识别原始高维电气信息中的重要电气特征(如故障场景下的重要机组), 进而实现算法的可解释性。

2) 因为本文采用发电机节点信息构建特征集, 算法天然具备统一的数据维度。在面向其他工作, 如需要处理不同维度的数据, 则在  $1 \times 1$  卷积的基础上进行局部卷积(local cov)<sup>[32]</sup>操作统一数据的特征维度, 进而实现注意力运算。

Transformer 编码器用来实现电气信息的分层表达, 提取有价值的高阶特征。本文在深层网络隐藏层内部加入概率为 50% 的 Dropout, 以削弱神经元间的共适应关系, 降低模型对特定特征的依赖, 避免模型过拟合。

本文构建的基于 Transformer 编码器的暂态评估判别模型可以直接面向底层数据, 建立量测数据与稳定类别之间的映射关系, 如图 2 所示。模型内部通过注意力运算, 自适应聚焦于高价值电气特征。

## 2.2 基于 Transformer 的多阶段暂稳评估流程

本文以电力系统量测数据为原始信息, 提出以 Transformer 分类模型为核心的多阶段暂态稳定评估流程, 包括离线训练和在线评估 2 个部分, 如图 3 所示。

1) 时域仿真阶段。

获取电力系统运行历史稳态数据和故障信息, 对电网不同运行方式时域暂稳仿真。稳定判据采用暂态稳定指数(transient stability index, TSI):

$$T_{SI} = \frac{360^\circ - |\Delta\delta_{\max}|}{360^\circ + |\Delta\delta_{\max}|} \quad (8)$$

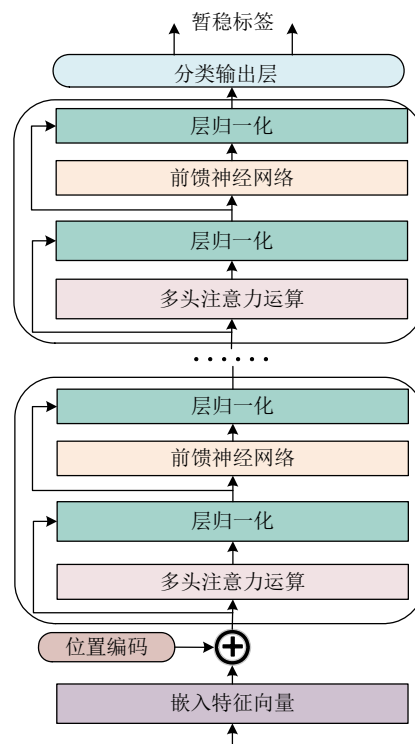


图 2 基于 Transformer 的 TSA 模型

Fig. 2 TSA model based on Transformer

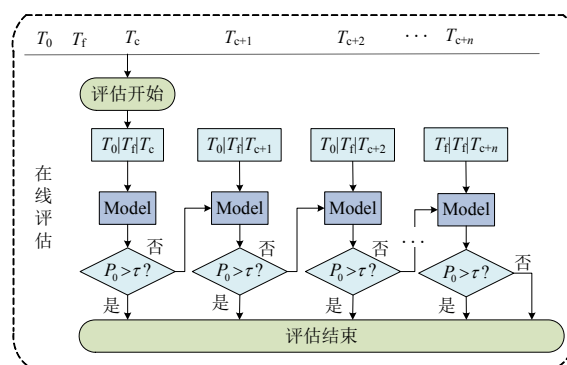
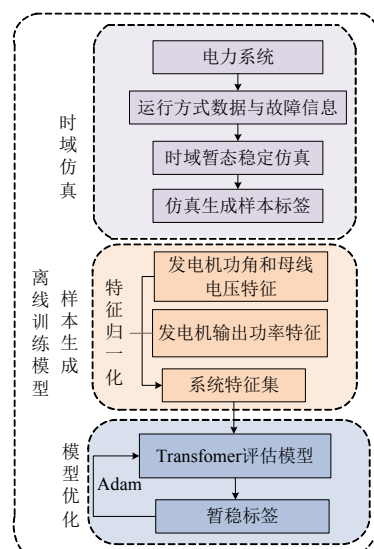


图 3 基于 Transformer 机制的多阶段暂态稳定评估流程

Fig. 3 Flowchart of the multi stage Transformer-based transient stability assessment

式中： $\Delta\delta_{\max}$  为仿真时长内任意 2 台发电机相对功角差的最大值。当  $T_{SI}>0$  时，样本判定为稳定，反之，判定为失稳。

2) 样本生成阶段。

电力系统暂态稳定性刻画了大扰动下电力系统的动态行为，相比采用单时刻电气信息，多断面电气量给自注意力点乘运算提供了更丰富的上下文信息，因此本文采用系统多时刻断面信息构成特征集。

为使样本集具备时空上的完备性，采用故障前稳态时刻  $t_0$ 、故障发生时刻  $t_f$  和故障清除后当前时刻  $t_{c+n}$  的信息构建特征空间， $t_0$  时刻信息采用故障发生前一周期的数据，样本采样频率 50Hz，具体见表 1。

表 1 模型特征集  
Table 1 Measurement features set

| 特征                                         |
|--------------------------------------------|
| $t_0$ 、 $t_f$ 、 $t_{c+n}$ 时刻，发电机母线电压       |
| $t_0$ 、 $t_f$ 、 $t_{c+n}$ 时刻，发电机有功、无功输出和功角 |

对输入矩阵采用 z-score 标准化处理，以消除属性量纲的差异影响，如式(9)所示，即

$$l' = \frac{l - \bar{L}}{\sigma_L} \tag{9}$$

式中： $\bar{L}$ 、 $\sigma_L$  为某一特征  $l$  的均值和标准差； $l'$  为归一化之后的值。

3) 模型优化阶段。

模型训练过程中通过降低整体的损失函数值来提高模型预测精度。本文采用交叉熵损失函数作为代价函数，因算法的复杂度较高对损失函数进行 L2 正则化，通过 L2 范数约束权重，防止模型对高维电气数据的过拟合：

$$L = -y_s \lg \hat{y} - (1 - y_s) \lg(1 - \hat{y}) + \frac{\lambda}{2} PW_{\text{weights}} P_2^2 \tag{10}$$

式中： $y_s$  为真实稳定样本； $\hat{y}$  为被预测为稳定样本的概率； $W_{\text{weights}}$  表示网络中的所有权重参数； $\lambda$  为 L2 正则化项的惩罚因子。

此外，为减少模型训练时间，本文采用 Adam 优化器自适应调节学习率，在损失函数梯度的相反方向上更新模型参数。

4) 在线应用阶段。

模型在故障清除时刻首次评估系统暂稳状态，以减少因故障清除初期暂态失稳特性不明显而导致的失稳漏判，每次评估间隔 0.02s，与样本生成阶段的采样频率统一。在故障清除时刻( $t_c$ )，PMU 实

时读取  $t_c$  量测数据，将稳态运行  $t_0$ 、故障发生时刻  $t_f$ 、故障清除时刻  $t_c$  的电气信息送入评估模型输出第一次暂稳判断。 $P_0$  为分类层 Softmax 函数输出的失稳概率<sup>[25]</sup>，当  $P_0>0.5$ ，输出失稳标签 0，反之，输出标签 1。为减少稳定场景的误判，引入失稳阈值  $\tau(\tau\geq0.5)$ ，若  $P_0>\tau$ ，判定为暂态失稳，由电网控制中心处理，如不满足，则继续读取( $t_c+1$ )时刻的信息，将  $t_0$ 、 $t_f$ 、 $t_c+1$  时刻的电气信息送入评估模型继续评估，在“第  $n$  次持续判定为稳定且最后 1s 内量测的发电机最大功角差均小于  $5^\circ$ <sup>[33]</sup>”后评估结束，此时系统恢复稳态。 $n$  根据实际电网需求和离线训练样本分布设定。本文中  $n$  设定为 750，即故障清除后 15s 内持续判稳且最后 1s 内发电机最大功角差小于  $5^\circ$  后模型退出运行。

3 仿真分析

3.1 数据集生成

算例采用基频为 50Hz 的 IEEE-39 节点电力系统。在 80%~120%的负荷水平内随机生成原始负荷样本，相应改变发电机出力使系统功率平衡且母线电压维持在 0.95~1.05pu 的合理范围。故障均发生于输电线路路上，故障地点位于距离线路首端的 2%、50%、98%处<sup>[27]</sup>。故障类型为 3 相短路，假设故障清除后网络拓扑结构不变，故障发生时刻为 0.02s，故障持续时间分别考虑 0.10s、0.15s、0.20s，仿真时长 20s。仿真软件采用 MATLAB power system tool (PST)3.0<sup>[34]</sup>，共仿真 18400 个场景，其中稳定场景 11580 个，失稳场景 6820 个。根据 8:1:1 比例随机划分训练集、验证集和测试集。

3.2 模型性能的评估指标

电力系统暂态稳定评估属于非平衡分类问题。仅采用评估准确率  $A_c$  作为评估模型的评价指标，会给模型评价带来误导。

为全面评估所提方法的性能，本文引入混淆矩阵，针对电气数据集合中稳定样本多于失稳样本的特点相应修改精确度和召回率指标，各类指标定义如表 2 所示。

表 2 暂态稳定评估混淆矩阵  
Table 2 Confusion matrix of TSA

| 评估模型 | 真实结果 |    |
|------|------|----|
|      | 稳定   | 失稳 |
| 稳定   | TP   | FP |
| 失稳   | FN   | TN |

$$A_c = \frac{T_p + T_N}{T_p + T_N + F_p + F_N} \quad (11)$$

$$F_1 = \frac{2p_{\text{recision}} \times r_{\text{ecall}}}{p_{\text{recision}} + r_{\text{ecall}}} \quad (12)$$

$$r_{\text{ecall}} = T_N / (T_N + F_p) \quad (13)$$

$$p_{\text{recision}} = T_N / (T_N + F_N) \quad (14)$$

式中：召回率  $r_{\text{ecall}}$  和精确率  $p_{\text{recision}}$  分别代表被正确预测的失稳样本占真实失稳样本和预测失稳样本的比例； $A_c$  反映评估模型的全局正确率； $F_1$  是  $r_{\text{ecall}}$  和  $p_{\text{recision}}$  的调和平均值，能更为客观地反映模型对失稳样本的预测能力。因此本文主要采取准确率  $A_c$  和  $F_1$  值评价模型性能。

### 3.3 模型评估性能

目前，时序特征抽取网络主要存在 RNN、CNN 和 Transformer 这 3 类。图 4 展示不同网络中时序提取结构，RNN 中的循环神经元通过记忆前序信息影响后序节点的输出；CNN 通过卷积核逐层提取局部以最终获取全局信息；Transformer 在位置编码层标记特征的相对位置，通过自注意力机制将实现全局特征的快速交互。

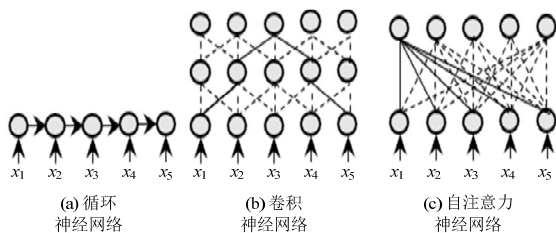


图 4 不同时序网络结构

Fig. 4 Architectures of different sequential neural networks

因此本文选取长短期记忆网络(long short-term memory, LSTM)和 CNN 作为深层对比模型，LSTM 是特殊结构的 RNN，具有与传统 RNN 相同的序列串行输入，但能有效缓解梯度弥散。同时采用暂稳分析领域常见模型 SVM 和 DT 作为对比算法。

本文所有算法的 Batchsize 为 400，深层模型均采用 epoch 最大 200，初始学习率 0.0025，L2 的惩罚因子 0.001，dropout 0.5，采用网格搜寻最优超参数。SVM 采用径向基函数， $C=10$ ， $\gamma=0.12$ ；DT 采用 C4.5 算法，置信因子为 0.2；Transformer 中  $h$  为 3，Transformer 子模块数为 2；LSTM 基于文献[35]中的结构，隐层数目 5，隐层单元的数目分别为 500, 600, 700, 600, 500；CNN 模型基于文献[16]中的 4 层卷积结构，卷积核  $3 \times 3$ ，每层卷积层后连接  $2 \times 2$

的最大池化层，以及 2 个全连接层构成。每层卷积核分别为 32, 32, 64, 64，激活函数 Relu。SVM 和 DT 不具备记录特征时序的能力，其特征需要相应调整为 1-D 维度。3.3~3.5 节检验各个对比算法的样本分类性能，所有模型在  $P_0 > 0.5$ ，即  $l_{\text{abel}}=0$  (失稳标签)时，将该场景判定为失稳。3.6 节讨论失稳阈值  $\tau$  对于多阶段模型漏判和误判率的影响， $\tau$  在降低误判的同时，可能导致模型对失稳样本识别时刻的延时，增加漏判风险，因此  $\tau$  需要根据模型权衡取值， $\tau$  取值不会影响模型训练过程。对于各个模型均进行 5 次重复实验，准确率取测试平均值。本文对比模型评估性能均采用多阶段预测后得出，深层模型的耗时对比见附表 A1。模型训练与测试均采用 Python 语言、基于 PyTorch 框架实现，服务器为 6240R 2.4G，24C/48T，数据存储为 .npz 格式。

因为失稳样本中的失稳特性会随着时间更加明显，使其更容易被模型正确识别，实验中我们发现，测试集所有失稳场景中系统最快在故障清除后 0.54s 失稳，为给信号传输和紧急控制的保留一定时间裕度，因此其失稳场景下的准确率定义为失稳样本是否能在失稳发生时刻(即  $T_{\text{SI}}=0$  的时刻)前 0.5s<sup>[36]</sup> 被模型正确识别，如表 3 所示。

表 3 Transformer 模型多阶段性能

Table 3 Performance in transformer multi-stage model

| 评估时刻      | 漏判个数 |
|-----------|------|
| $t_c$     | 12.2 |
| $t_{c+1}$ | 7.8  |
| $t_{c+2}$ | 7.6  |
| $t_{c+3}$ | 7.0  |
| $t_{c+4}$ | 5.2  |

通过在测试集中 1840 个场景(其中失稳场景 892 个)进行预测，漏判个数取 5 次重复试验下平均值。表 3 验证多阶段模型能够有效降低失稳场景漏判，在第一次预判时，模型平均漏判样本 12.2 个。98.6% 的失稳样本被识别。因机组的失稳特性会随时间愈发明显，远离模型分类边界。截止到失稳发生(即  $T_{\text{SI}}=0$  的时刻)前 0.5s，漏判个数为 1.2 个，此时模型准确率 99.53%， $F_1$  值 99.48。

表 4 展示多种算法下多阶段预测最终的准确率和  $F_1$  值，Transformer 的  $F_1$  值分别比 CNN 和 LSTM 高出 0.48% 和 0.24%。3 种深层模型在采用多阶段预测后，其准确率均高于 99%，相比常见的单阶段模型<sup>[10-16]</sup>，在准确率上有一定提升。

同时，对比现有文献中常见的“特征提取+浅

表 4 不同模型下的模型性能比较  
Table 4 Performance comparison  
between different models

| 评估模型        | 评估准确率 $A_c/\%$ | $F_1/\%$ |
|-------------|----------------|----------|
| Transformer | 99.53          | 99.48    |
| CNN         | 99.28          | 99.00    |
| LSTM        | 99.34          | 99.24    |
| PCA+SVM     | 96.68          | 95.81    |
| PCA+DT      | 95.91          | 95.09    |

层算法”。从上述数据可以看出，Transformer 模型性能明显优于前置主成分分析法 (principal component analysis, PCA) 的 SVM 和 DT。表明本文模型相比 SVM 和 DT，在处理高维数据时具有更优的复杂函数表征能力。

3.4 模型中注意力分布及可视化

注意力机制因其良好的特征权重分布可视化，被逐渐应用于模型的可解释性分析。注意力的可解释性具备体现为：注意力权重的高低对应特征的重要程度，高权重的特征单元对输出结果具有更高决定性<sup>[20]</sup>。

在自然语言处理和计算机视觉领域，注意力机制实现了模型对关键词语<sup>[20]</sup>和关键像素<sup>[23]</sup>的关注，本节用以探讨注意力机制是否适用于电力系统这种复杂物理模型，进而分析电气特征与暂稳输出之

间的映射关系。如图 5 所示颜色越深代表该特征分配到的注意力权重越高，本节分别分析模型在面向时间特征和机组特征时的注意力分布，证明 Transformer 的注意力机制在面向电网物理特征时能够实现对高维电气时空特征的辨识。

3.4.1 面向时间维度时的注意力分布

图 5 为 100 个场景中模型对时间断面的注意力分布。随着训练次数增多，模型注意力概率明显向  $t_{c+n}$  时刻，即故障清除后的特征偏移。图 5(d)呈现出与图 5(c)相似的注意力分布，说明模型准确率在第 20 轮训练后已经达到了较高值，模型梯度更新缓慢，其注意力分布趋于稳定，统计第 100 轮训练时不同时刻特征的注意力分布，得出  $t_0$ 、 $t_f$ 、 $t_{c+n}$  时刻的注意力分布分别占 5.29%、24.03%和 70.68%。

上述结果表明，在时间维度上，故障清除后的数据对暂稳评估价值度最高，模型赋予更多的注意力权重，而稳态状态时的信息价值度最低。为验证电气特征间注意力分布的有效性，并说明采用  $t_0$ 、 $t_f$ 、 $t_{c+n}$  信息相比单一时刻数据作为特征子集能够提升模型性能。表 5 对比了模型在故障清除时刻  $t_c$  的首次评估准确率，即采用  $t_c$  时刻， $t_f$ 、 $t_c$  时刻和  $t_0$ 、 $t_f$ 、 $t_c$  时刻特征构建特征集分别训练模型。

由表 5 可知，一方面，仅采用  $t_c$  时刻的数据依

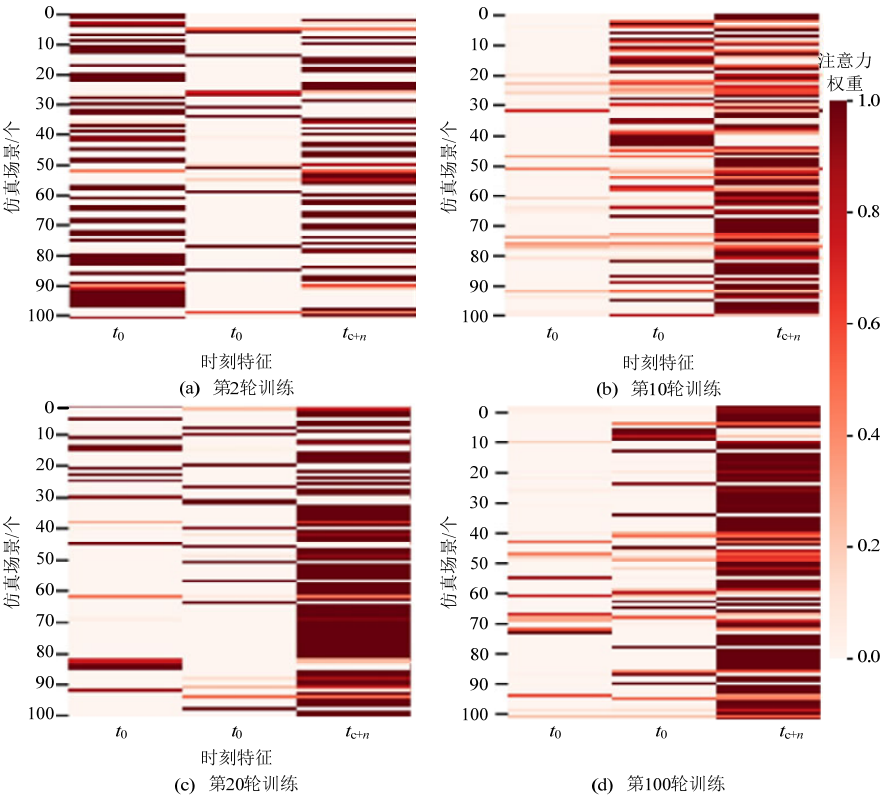


图 5 训练过程中不同时段特征的注意力分布

Fig. 5 Attention distribution for different temporal data during train process

表 5 采用不同时间断面特征后模型性能比较  
Table 5 Performance comparison of transformer with different temporal features

| 时间断面     | $t_c$ | $t_c \sim t_f$ | $t_0 \sim t_f \sim t_c$ |
|----------|-------|----------------|-------------------------|
| $F_1/\%$ | 95.99 | 96.84          | 98.71                   |

然能使模型  $F_1$  值达到 95.99% 的较高水平，说明注意力有效识别了高价值特征，与图 5 结果相符，同时表明高注意力权重的特征会更大程度上影响模型决策；另一方面，在采用全部断面特征时，评估性能会较单断面输入明显提升，由此推断，即使  $t_0$  时刻的注意力权重仅占 5.29%，但  $t_0$  时刻数据通过对全局特征的注意力融合运算，为模型注意力交互提供了有效信息，进而提升模型性能。

3.4.2 面向发电机组时的注意力分布

如果采用发电机电气信息作为模型 Embedding 层，模型中不再采用 Position Embedding 显式标注特征时序。如图 6、7 所示，模型对不同故障下的机组呈现了不同的注意力权重分布，证明模型在面对不同故障时能够自适应地更加关注于某些发电机的量测信息，而这些发电机的状态变化会更大程度上影响系统的暂稳状态。

图 6 展示了包含全部故障场景的注意力分布。

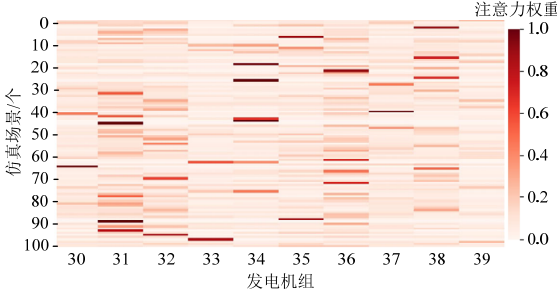


图 6 面向全部场景的发电机注意力分布  
Fig. 6 Generator attention distribution under different scenarios

因不同故障下的机组状态不同，机组注意力会呈现出相对均衡的分布。从原始注意力分布中提取故障发生于线路 6-5 和 20-34 中的注意力分布，如图 7 所示。

由图 7(a)可以看出，当不同负荷场景下，当故障发生于输电线 6-5 时，模型的注意力分布显著集中于 G39，在 100 组数据中的注意力平均值为 79.02%，明显高于其他机组。这表明 G39 在暂态稳定和失稳的不同场景下，其运行状态会有区别于其他机组。为验证上述结论，在 MATLAB 中进行故障时域仿真可知，在该故障下，如果发生暂态失稳，与其他机组失去同调性的机组为 G39，如图 7(a)

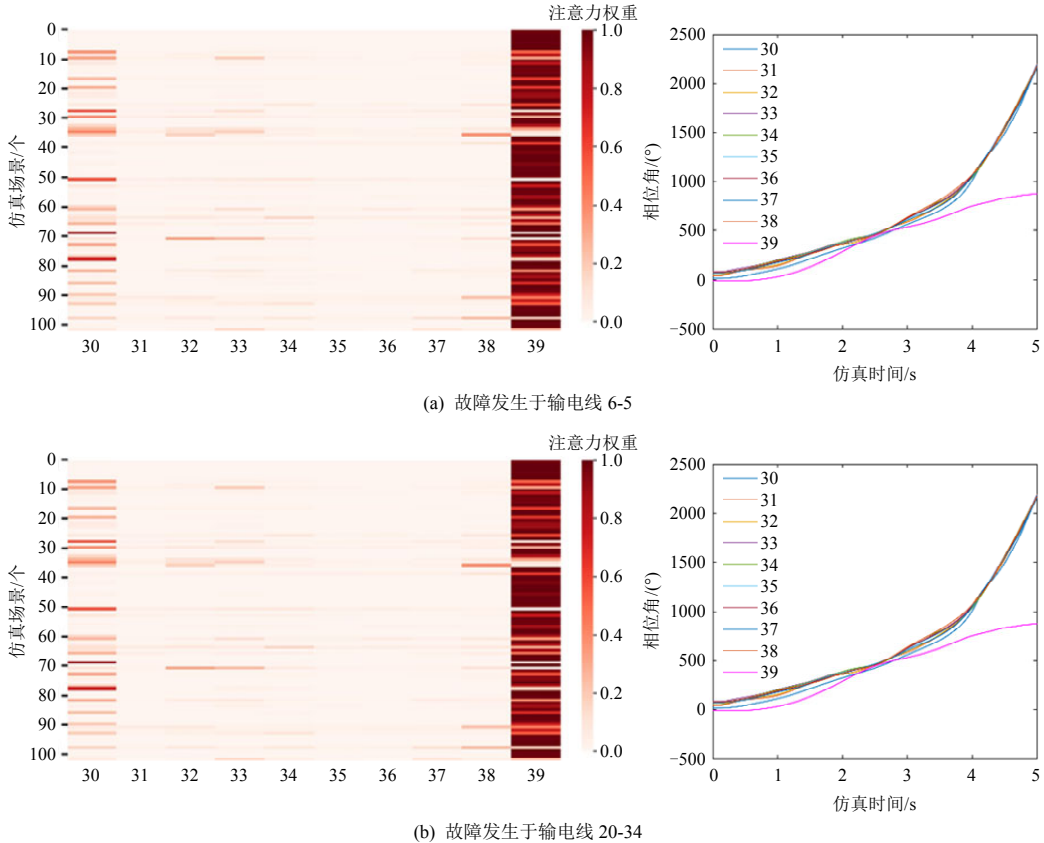


图 7 面向特定故障下的发电机注意力分布  
Fig. 7 Generator attention distribution under certain fault

所示。

同理,图 7(b)中模型注意力自适应聚焦于 G34 和 G39,注意力分布分别为 49.87%和 33.11%。从 MATLAB 仿真结果可知,当故障发生于输电线 20-34, G34 机组为主要失同步机组,而 G39 机组在本文仿真中为平衡机组, G39 通过调节有功和无功出力维持仿真系统的稳定,因此其特征变动相对较大,模型也会给其分配较高的注意力权重。统计同一故障场景下的注意力分布后发现,除平衡机组外,每个机组的注意力权重聚焦于该故障下的失稳机组,且同类故障场景具备相似的机组注意力分布。

从本节可以得出,Transformer 算法中实现了深

层网络对关键特征的自适应加权,识别出高价值的电气特征,并证明高注意力权重的特征会更大程度上影响模型决策,面向暂稳分析问题,Transformer 模型较其他常见深层网络具有了有效的可解释性。

### 3.5 模型面对异常特征的鲁棒性

因实际应用场景中可能存在数据污染,为验证注意力运算对异常特征的抑制能力,将样本集内的部分数据遮挡并替换为随机值。新加入随机特征向量服从高斯分布,均值为 0,方差为 1。

为展示 Transformer 模型内部对异常信息的处理,将  $t_{c+n}$  时刻的特征随机遮挡 30%,注意力可视化结果如图 8 所示。

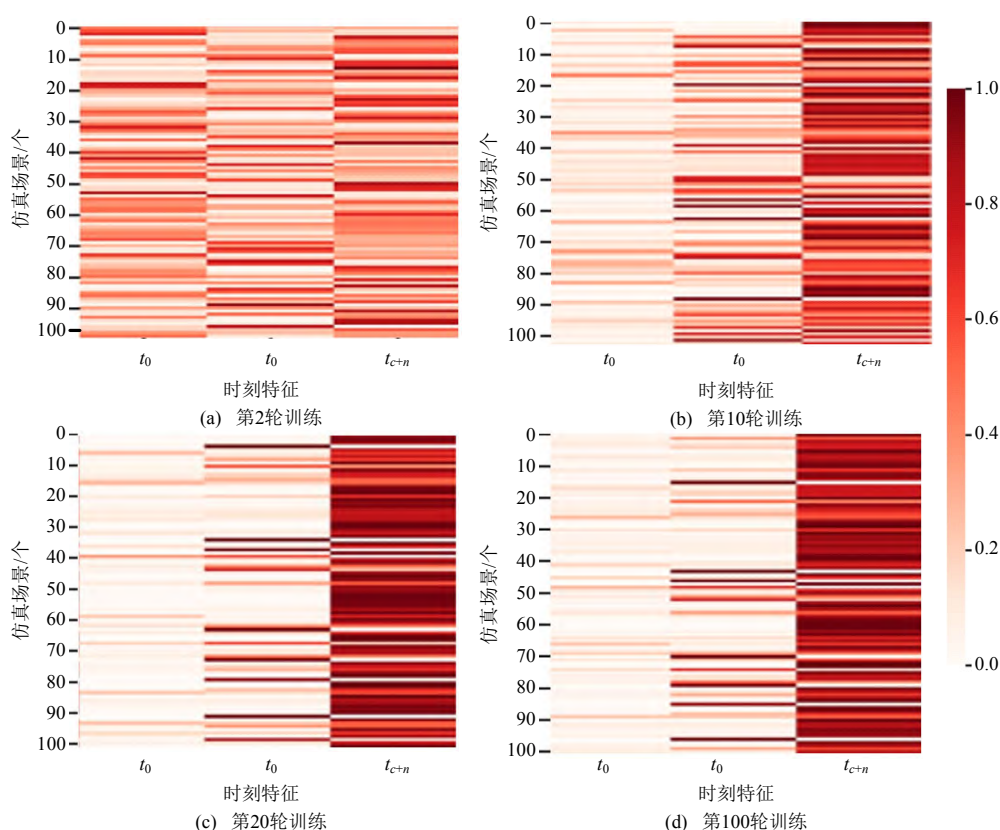


图 8 在特征遮挡下不同时段特征的注意力分布

Fig. 8 Attention distribution for different temporal data during train process with feature masked

大量遮挡最具价值的故障清除时刻后的特征会导致模型训练减缓。在第 2 轮训练中,不同特征之间的注意力差异并不明显。与之相对,在图 5(a)中,模型已经明显呈现出对不同特征的注意力倾斜。

然而随着迭代次数增加,模型最终依然聚焦于故障清除后时刻。在图 8(d)的第 100 轮训练中,注意力在  $t_0$ 、 $t_1$ 、 $t_{c+n}$  时刻的分布值为 9.21%、19.34%和 71.45%,与图 5(d)呈现出相近的注意力分布。这表明  $t_{c+n}$  时刻未被污染的数据依然会为模型判别提供最具价值的信息,模型关注于故障清除时刻。

对比不同模型对异常数据的处理能力,如表 6 所示,异常数据导致所有模型的性能有了不同程度的下降。其中,Transformer 表现出最强的鲁棒性,在异常数据占比 10%时准确率和  $F_1$  值仅下降 0.23%和 0.46%,当异常数据占比 30%时,依然能维持 95%以上的准确率和  $F_1$  值,高于其他模型。Transformer 特征间以点乘方式进行自注意力运算,其全局感受野使其具备了相比常见算法更强的上下文信息整合能力<sup>[29-31]</sup>,使模型在面对污染数据时维持较高的准确率。

表 6 含异常特征时模型的性能比较

Table 6 Performance comparison containing anomalous features

| 评估模型        | 评估准确率 $A_c/\%$      |                     |                    | $F_1/\%$           |                     |                     |
|-------------|---------------------|---------------------|--------------------|--------------------|---------------------|---------------------|
|             | 异常特<br>征占比<br>为 10% | 异常特<br>征占比<br>为 20% | 异常特征<br>占比为<br>30% | 异常特征<br>占比为<br>10% | 异常特<br>征占比<br>为 20% | 异常特<br>征占比<br>为 30% |
|             |                     |                     |                    |                    |                     |                     |
| Transformer | 99.30               | 98.32               | 96.09              | 99.02              | 97.92               | 95.17               |
| CNN         | 98.93               | 97.25               | 93.81              | 98.28              | 96.70               | 92.89               |
| LSTM        | 98.89               | 93.49               | 89.70              | 98.29              | 92.93               | 88.96               |
| PCA+SVM     | 95.59               | 93.03               | 91.95              | 94.74              | 92.17               | 90.67               |
| PCA+DT      | 94.88               | 92.86               | 89.97              | 94.20              | 92.33               | 88.94               |

在含有大量异常数据的场景下，CNN 模型较 LSTM 表现出了更好的性能，因其池化层中的信息降维削弱了无效信息对模型的干扰。LSTM 模型对特征污染表现出很高的敏感度，随着异常数据的增多评估性能急速下降，在异常数据 30% 时，LSTM 的准确率和  $F_1$  值下降到 90% 以下，低于 PCA+SVM 网络，这表明 LSTM 性能依赖于输入信息的完整度，不适用于可能存在大量数据污染的应用场景。

### 3.6 模型失稳阈值的 $\tau$ 分析

本文采用 Softmax 函数<sup>[37]</sup>作为分类层函数，分类层对每一组样本输出失稳概率  $P_0$  和稳定概率  $P_1$ ，二者相加为 100%。当  $P_1 > 0.5$  时， $P_1 > P_0$ ，模型  $l_{\text{label}}=1$ 。 $P$  可以视为标签的可信程度。

为降低模型的稳定样本误判率，模型输出失稳判定的条件由上文中“ $P_0 > 0.5$ ”即“ $l_{\text{label}}=0$ ”，进一步限制为“ $P_0 > \tau (\tau \geq 0.5)$ ”。验证集试验结果如图 9 所示。

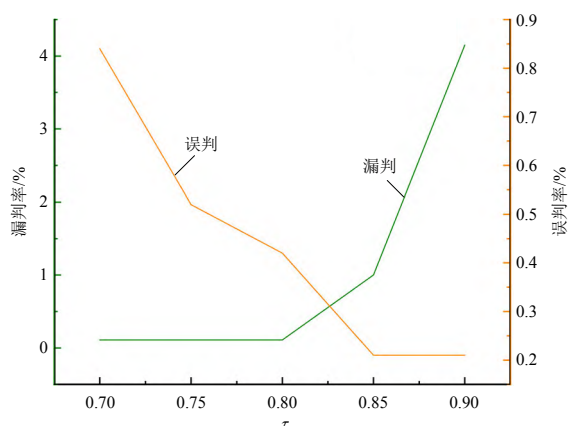


图 9 不同失稳阈值下的模型性能

Fig. 9 Performance comparison under different instability criterion

试验结果表明，失稳阈值  $\tau$  的加入会降低误判率。为使模型准确率更具参考性，本文加入了“失稳样本需要在失稳发生时刻前 0.5s 识别，否则判定

为漏判”这一限制，由于较高的稳定失稳阈值会造成失稳识别时刻的延后，因此  $\tau$  的加入会在降低误判率的同时，使模型漏判率有一定升高，当  $\tau$  大于 0.80，漏判率升高明显。因此将  $\tau$  设定为 0.80，测试集结果如下，在  $\tau$  由 0.5 升为 0.80 时，漏判率由 0.13% 升为 0.16%，误判率由 0.84% 降为 0.42%，模型最终准确率 99.70%， $F_1$  为 99.69%。

本文采用多阶段的暂态评估方法，在故障清除时刻首次输出暂稳评估。因失稳特性随着时间推移愈发明显，模型漏判率降低。如附表 A2 所示，3 种算法均能够达到较高的准确率，相比现有文献，本文加入“在失稳发生时刻前 0.5s 识别出失稳”这一限制，如果模型在失稳时刻前 0.5s 内没有识别出失稳，则被判定为漏判，给紧急控制保留一定时间，因此本文的漏判率更具实际参考性。

本文方法具有以下特点：1) 特征维度较低，不需要人工特征筛选，采用稳态、故障发生和清除时刻信息构建的特征空间给算法提供了丰富的前后文信息；2) 相比 CNN 和 DBN，Transformer 不仅能够提取时序关系，还能够快速提取全局信息，表现出良好的分类性能；3) 自注意力结构带来了良好的可解释性，一方面给出特征权重，为数据收集提供方向，另一方面将深层网络透明化，增加了神经网络在实际应用中的可信性。

## 4 结论

1) 本文构建的多阶段评估模型面向底层量测数据预测系统暂稳状态，其点乘自注意力机制实现了全局特征间的快速交互，并充分利用电力系统时序信息，其性能高于常见算法和单次预测模型。

2) 本文方法展现出良好的可解释性，展现深层网络内部工作过程自注意力引导模型在训练迭代中自适应聚焦于重要特征，揭示不同特征对于决策的贡献度，提高了算法决策的可信度，并给未来的数据收集提供了方向。并通过验证注意力分布的合理性和高权重特征对模型的更大影响，证明注意力机制在电力系统暂稳分析中的有效性。

3) 本文方法具有性能优越的鲁棒性，本文证明了注意力机制有助于评估模型在污染数据下维持较高性能，表明本文模型在真实电力系统中的适用性。

本文后续的研究在于基于本文 Transformer 模型拟合系统紧急控制切机容量，优化切机方案，进

一步提高模型的实用性。

## 参考文献

- [1] 李亚楼, 李芳, 刘赫川, 等. 基于 PSASP 的综合能源仿真分析系统[J]. 电网技术, 2019, 43(7): 2498-2505.  
LI Yalou, LI Fang, LIU Hechuan, et al. Framework design of simulation system for integrated energy systems based on PSASP[J]. Power System Technology, 2019, 43(7): 2498-2505(in Chinese).
- [2] 薛禹胜. EEAC 与直接法的机理比较(一)——受扰程度函数[J]. 电力系统自动化, 2001, 25(11): 6-11.  
XUE Yusheng. A critical comparison of various methods for transient stability assessment part one: disturbed-measure functions[J]. Automation of Electric Power Systems, 2001, 25(11): 6-11(in Chinese).
- [3] SAMANTARAY S R, KAMWA I, JOOS G. Phasor measurement unit based wide-area monitoring and information sharing between micro-grids[J]. IET Generation, Transmission & Distribution, 2017, 11(5): 1293-1302.
- [4] 孙宏斌, 王康, 张伯明, 等. 采用线性决策树的暂态稳定规则提取[J]. 中国电机工程学报, 2011, 31(34): 61-67.  
SUN Hongbin, WANG Kang, ZHANG Boming, et al. Rule extraction in transient stability study using linear decision trees[J]. Proceedings of the CSEE, 2011, 31(34): 61-67(in Chinese).
- [5] 马骞, 杨以涵, 刘文颖, 等. 多输入特征融合的组合支持向量机电力系统暂态稳定评估[J]. 中国电机工程学报, 2005, 25(6): 17-23.  
MA Qian, YANG Yihan, LIU Wenying, et al. Power system transient stability assessment with combined SVM method mixing multiple input features[J]. Proceedings of the CSEE, 2005, 25(6): 17-23(in Chinese).
- [6] 叶圣永, 王晓茹, 刘志刚, 等. 基于支持向量机增量学习的电力系统暂态稳定评估[J]. 电力系统自动化, 2011, 35(11): 15-19.  
YE Shengyong, WANG Xiaoru, LIU Zhigang, et al. Power system transient stability assessment based on support vector machine incremental learning method [J]. Automation of Electric Power Systems, 2011, 35(11): 15-19.
- [7] 叶圣永, 王晓茹, 刘志刚, 等. 基于随机森林算法的电力系统暂态稳定性评估[J]. 西南交通大学学报, 2008, 43(5): 573-577.  
YE Shengyong, WANG Xiaoru, LIU Zhigang, et al. Transient stability assessment based on random forest algorithm[J]. Journal of Southwest Jiaotong University, 2008, 43(5): 573-577(in Chinese).
- [8] SCHMIDHUBER J. Deep learning in neural networks: an overview[J]. Neural Networks, 2015, 61: 85-117.
- [9] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [10] 朱乔木, 党杰, 陈金富, 等. 基于深度置信网络的电力系统暂态稳定评估方法[J]. 中国电机工程学报, 2018, 38(3): 735-743.  
ZHU Qiaomu, DANG Jie, CHEN Jinfu, et al. A method for power system transient stability assessment based on deep belief networks[J]. Proceedings of the CSEE, 2018, 38(3): 735-743(in Chinese).
- [11] 胡伟, 郑乐, 闵勇, 等. 基于深度学习的电力系统故障后暂态稳定评估研究[J]. 电网技术, 2017, 41(10): 3140-3146.  
HU Wei, ZHENG Le, MIN Yong, et al. Research on power system transient stability assessment based on deep learning of big data technique[J]. Power System Technology, 2017, 41(10): 3140-3146(in Chinese).
- [12] 王怀远, 陈启凡. 基于代价敏感堆叠变分自动编码器的暂态稳定评估方法[J]. 中国电机工程学报, 2020, 40(7): 2213-2220.  
WANG Huaiyuan, CHEN Qifan. A transient stability assessment method based on cost-sensitive stacked variational auto-encoder[J]. Proceedings of the CSEE, 2020, 40(7): 2213-2220(in Chinese).
- [13] 朱乔木, 陈金富, 李弘毅, 等. 基于堆叠自动编码器的电力系统暂态稳定评估[J]. 中国电机工程学报, 2018, 38(10): 2937-2946.  
ZHU Qiaomu, CHEN Jinfu, LI Hongyi, et al. Transient stability assessment based on stacked autoencoder [J]. Proceedings of the CSEE, 2018, 38(10): 2937-2946(in Chinese).
- [14] TAN Bendong, YANG Jun, PAN Xueli, et al. Representational learning approach for power system transient stability assessment based on convolutional neural network[J]. The Journal of Engineering, 2017, 2017(13): 1847-1850.
- [15] BALASUBRAMANIAM K, VENAYAGAMOORTHY G K, WATSON N. CNN based power system transient stability margin and voltage stability index prediction[C]// 2013 IEEE Computational Intelligence Applications in Smart Grid(CIASG). Singapore: IEEE, 2013: 13-20.
- [16] 高昆仑, 杨帅, 刘思言, 等. 基于一维卷积神经网络的电力系统暂态稳定评估[J]. 电力系统自动化, 2019, 43(12): 18-26.

- GAO Kunlun, YANG Shuai, LIU Siyan, et al. Transient stability assessment for power system based on one-dimensional convolutional neural network[J]. Automation of Electric Power Systems, 2019, 43(12): 18-26(in Chinese).
- [17] YU J J Q, HILL D J, LAM A Y S, et al. Intelligent time-adaptive transient stability assessment system[J]. IEEE Transactions on Power Systems, 2018, 33(1): 1049-1058.
- [18] BALTAS N G, MAZIDI P, MA Jin, et al. A comparative analysis of decision trees, support vector machines and artificial neural networks for on-line transient stability assessment[C]//2018 International Conference on Smart Energy Systems and Technologies(SEST). Seville: IEEE, 2018: 1-6.
- [19] EPELBAUM T. Deep learning: Technical introduction [Z]. arXiv: 1709.01412, 2017.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 6000-6010.
- [21] ZHAO Qinlu, CAI Xiaodong, CHEN Chaocun, et al. Commented content classification with deep neural network based on attention mechanism[C]//Proceedings of the 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference. Chongqing: IEEE, 2017: 2016-2019.
- [22] YANG Baosong, TU Zhaopeng, WONG D F, et al. Modeling localness for self-attention networks[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018: 4449-4458.
- [23] RANFTL R, BOCHKOVSKIY A, KOLTUN V. Vision transformers for dense prediction[C]//2021 IEEE/CVF International Conference on Computer Vision(ICCV). Montreal: IEEE, 2021.
- [24] LIU Xiaodong, DUH K, LIU Liyuan, et al. Very deep Transformers for neural machine translation [Z]. arXiv:2008.07772v2, 2020.
- [25] LUNDBERG S M, ERION G, CHEN H, et al. From local explanations to global understanding with explainable AI for trees[J]. Nature Machine Intelligence, 2020, 2(1): 56-67.
- [26] 邵美阳, 吴俊勇, 李宝琴, 等. 基于两阶段集成深度置信网络的电力系统暂态稳定评估[J]. 电网技术, 2020, 44(5): 1776-1787.
- SHAO Meiyang, WU Junyong, LI Baoqin, et al. Transient stability assessment of power system based on two-stage ensemble deep belief network[J]. Power System Technology, 2020, 44(5): 1776-1787(in Chinese).
- [27] 吴俊勇, 张若愚, 季佳仲, 等. 计及漏判/误判代价的两阶段电力系统暂态稳定预测方法[J]. 电力系统自动化, 2020, 44(24): 44-52.
- WU Junyong, ZHANG Ruoyu, JI Jiashen, et al. Two-stage transient stability prediction method of power system considering cost of misdetection and false alarm [J]. Automation of Electric Power Systems, 2020, 44(24): 44-52(in Chinese).
- [28] 李福成, 徐箭, 廖思阳, 等. 基于样本关注度和多层次特征的多阶段电力系统暂态稳定评估[J]. 中国电机工程学报, 2021, 41(22): 7596-7608.
- LI Fucheng, XU Jian, LIAO Siyang, et al. Multi-stage power system transient stability assessment based on sample attention and hierarchical feature[J]. Proceedings of CSEE, 2021, 41(22): 7596-7608(in Chinese).
- [29] WANG Xiaolong, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018.
- [30] GUO Yu, GOU Gaopeng, XIONG Gang, et al. Let imbalance have nowhere to hide: Class-sensitive feature extraction for imbalanced traffic classification[C]//2021 International Joint Conference on Neural Networks (IJCNN). Shenzhen: IEEE, 2021.
- [31] NASEER M, RANASINGHE K, KHAN S, et al. Intriguing properties of vision transformers[J]. 2021, arXiv:2105.10497v1.
- [32] GREGOR K, LECUN Y. Emergence of complex-like cells in a temporal product network with local receptive fields [Z]. arXiv: 1006.0448v1, 2010.
- [33] SHRESTHA B, GOKARAJU R R, SACHDEV M. Out-of-step protection using state plane trajectories analysis [J]. IEEE Transactions on Power Delivery, 2013, 28(2): 1083-1093.
- [34] CHOW J H, CHEUNG K W. A toolbox for power system dynamics and control engineering education and research [J]. IEEE Transactions on Power Systems, 1992, 7(4): 1559-1564.
- [35] 孙黎霞, 白景涛, 周照宇, 等. 基于双向长短期记忆网络的电力系统暂态稳定评估[J]. 电力系统自动化, 2020, 44(13): 64-72.
- SUN Lixia, BAI Jingtao, ZHOU Zhaoyu, et al. Transient stability assessment of power system based on bi-

directional long-short-term memory network[J] . Automation of Electric Power Systems, 2020, 44(13): 64-72(in Chinese).

[36] Gomez F R, RAJAPAKSE A D, ANNAKKAGE U D, et al . Support vector machine-based algorithm for post-fault transient stability status prediction using 附录 A

synchronized measurements[J]. IEEE Transactions on Power Systems, 2011, 26(3): 1474-1483.

[37] WASEF M, RAFLA N. Hardware implementation of Multi-Rate input SoftMax activation function[C]//2021 IEEE International Midwest Symposium on Circuits and Systems(MWSCAS). Lansing: IEEE, 2021: 783-786.

表 A1 深层模型耗时比较  
Table A1 Performance comparison between deep models

| 评估模型        | 训练 50epoch 耗时/s | 在线评估耗时/s |
|-------------|-----------------|----------|
| Transformer | 1197.65         | 0.015    |
| CNN         | 702.02          | 0.011    |
| LSTM        | 798.24          | 0.019    |

表 A2 模型与现有成果对比  
Table A2 Comparison between the proposed model and current works

| 评估模型   | 算法          | 特征选择                                     | 样本量   | 该文献中<br>最终准确率/%            | 漏判率/%                    | 误判率/%                    | 是否具备<br>可解释性 |
|--------|-------------|------------------------------------------|-------|----------------------------|--------------------------|--------------------------|--------------|
| 文献[25] | DBN         | 原始特征(功角、有功、无功等)、<br>人工经验特征和 SDAE 提取特征的集成 | 37400 | 99.32                      | 没有写明                     | 没有写明                     | 否            |
| 文献[26] | CNN         | 功角、电压幅值、功角轨迹簇、<br>电压幅值轨迹簇                | 37400 | 99.68<br>(累计到<br>第 10 次检测) | 0.12<br>(累计到<br>第 6 次检测) | 0.32<br>(累计到<br>第 6 次检测) | 否            |
| 本文     | Transformer | 功角、电压幅值、有功、无功                            | 18400 | 99.70                      | 0.16                     | 0.42                     | 是            |



房佳姝

在线出版日期: 2022-08-31。  
收稿日期: 2021-12-29。  
作者简介:  
房佳姝(1993), 女, 博士研究生, 研究方向为人工智能在电力系统中的应用, jiashu.fang@hotmail.com;  
刘崇茹(1977), 女, 博士, 教授, 研究方向为交直流混合输电系统运行与控制, chongru.liu@ncepu.edu.cn;  
苏晨博(1991), 男, 博士, 研究方向为新能源发电系统并网与控制, scb2636@foxmail.com;  
林晗星(1998), 男, 硕士, 研究方向为机器学习在电力系统中的应用, linhanxing19@163.com;  
\*通信作者: 郑乐(1989), 男, 博士, 研究方向为人工智能在电力系统分析与控制中的应用, zhengl20@ncepu.edu.cn。

(责任编辑 乔宝榆)

# Multi-stage Transient Stability Assessment of Power System Based on Self-attention Transformer Encoder

FANG Jiashu, LIU Chongru, SU Chenbo, LIN Hanxing, ZHENG Le\*

(State Key Lab. for Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University))

**KEY WORDS:** power system; transient stability assessment (TSA); Transformer model; self-attention; attention visualization

The conventional deep networks for power system transient stability assessment (TSA) are generally considered "black box" models, limiting their practical applicability. Furthermore, conventional methods have limited ability to capture the dynamic evolution process of the power system. To solve the above problems, this paper introduces a multi-stage transient stability assessment method based on Transformer encoder. The Transformer model presents good interpretability, whose attention structure visualizes the internal work of the deep neural networks, so the model can adaptively identify and focus on the important features. Meanwhile, Transformer gains a global receptive field via attention mechanism, and thus the state change of power system can be characterized quickly and accurately.

As seen in Fig. 1, the information collected in fault clearance moment ( $t_c$ ), fault occurrence moment ( $t_f$ ), and pre-fault ( $t_0$ ) send into assessment model, the TSA would be predicted as being unstable when the unstable probability ( $P_0$ ) is higher than the unstable criterion ( $\tau$ ). The comparison of model performance across different data scenarios is presented in Tables 1 and 2.

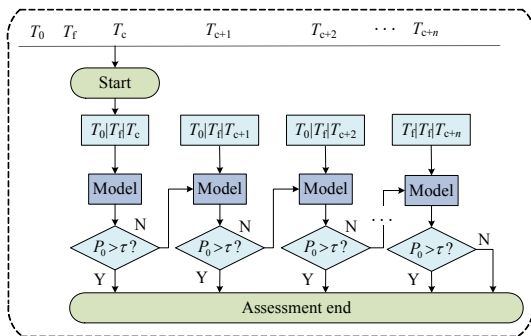


Fig. 1 Flowchart of the multi-stage online assessment

Table 1 Performance comparison between different models

| Model       | Accuracy/% | $F_1$ /% |
|-------------|------------|----------|
| Transformer | 99.53      | 99.48    |
| CNN         | 99.28      | 99.00    |
| LSTM        | 99.34      | 99.24    |
| PCA+SVM     | 96.68      | 95.81    |
| PCA+DT      | 95.91      | 95.09    |

Table 2 Performance comparison with anomalous data

| Model       | Accuracy/%                       |                                  |                                  | $F_1$ /%                         |                                  |                                  |
|-------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
|             | Proportion of anomalous data 10% | Proportion of anomalous data 20% | Proportion of anomalous data 30% | Proportion of anomalous data 10% | Proportion of anomalous data 20% | Proportion of anomalous data 30% |
|             | data 10%                         | data 20%                         | data 30%                         | data 10%                         | data 20%                         | data 30%                         |
| Transformer | 99.30                            | 98.32                            | 96.09                            | 99.02                            | 97.92                            | 95.17                            |
| CNN         | 98.93                            | 97.25                            | 93.81                            | 98.28                            | 96.70                            | 92.89                            |
| LSTM        | 98.89                            | 93.49                            | 89.70                            | 98.29                            | 92.93                            | 88.96                            |
| PCA+SVM     | 95.59                            | 93.03                            | 91.95                            | 94.74                            | 92.17                            | 90.67                            |
| PCA+DT      | 94.88                            | 92.86                            | 89.97                            | 94.20                            | 92.33                            | 88.94                            |

The Transformer model presents better performance compared with existing methods, and the final accuracy via multi-stage assessment reaches 99.53%.

Anomalous data causes performance degradation among difficult models. The transformer shows the strongest robustness. With the 30% anomalous data proportion, the proposed method still maintains accuracy higher than 95%.

The scaled dot-product attention layer is the core component of the Transformer model. The scaled dot-product attention is calculated by:

$$A_{\text{attention}}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = S_{\text{softmax}}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

where query ( $\mathbf{Q}$ ), key ( $\mathbf{K}$ ), value ( $\mathbf{V}$ ) are three different learnable matrices linearly transformed from the input sequence.  $d_k$  is a scaling factor. The attention distribution determines the relative importance of feature tokens in the whole sequence.

The attention distribution can focus on the important spatiotemporal features under high-dimensional electrical information. Specifically, the information on fault clearance moment presents higher attention distribution relative to fault occurrence and pre-fault moment. Besides, attention distribution tends to focus on the unstable generators and slack generator under difficult operations. Moreover, it can be clearly noticed that the attention distribution has important features during the training process.