

# 基于改进 LightGBM 的电力系统暂态稳定评估方法

周挺<sup>1</sup>, 杨军<sup>1</sup>, 周强明<sup>2</sup>, 谭本东<sup>1</sup>, 周悦<sup>2</sup>, 徐箭<sup>1</sup>, 孙元章<sup>1</sup>

(1. 武汉大学 电气与自动化学院, 湖北省 武汉市 430072;

2. 国网湖北省电力公司, 湖北省 武汉市 430077)

## Power System Transient Stability Assessment Method Based on Modified LightGBM

ZHOU Ting<sup>1</sup>, YANG Jun<sup>1</sup>, ZHOU Qiangming<sup>2</sup>, TAN Bendong<sup>1</sup>, ZHOU Yue<sup>2</sup>, XU Jian<sup>1</sup>, SUN Yuanzhang<sup>1</sup>

(1. School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, Hubei Province, China;

2. State Grid Hubei Electric Power Company, Wuhan 430077, Hubei Province, China)

**ABSTRACT:** The actual available operation data of power system generally contain the noise caused by the measurement error of wide area measurement system (WAMS), and are also characterized by class imbalance, leading to great influence on the classification performance of the power system transient stability evaluation algorithms based on machine learning. This paper proposes a transient stability assessment method based on modified LightGBM (Light Gradient Boosting Machine). This histogram algorithm is used for data discretization to strengthen the robustness to noise, and larger weights are given to unstable samples in training process to mitigate the impact of imbalanced samples. Meanwhile, a regularization term is introduced into the loss function to reduce overfitting. The simulation results on New England 10-machine 39-bus system and an actual 500-bus grid in State of South Carolina, USA show that, compared with other machine learning methods, the proposed model is hardly prone to over-fitting under noise interference with better robustness. It has a higher recognition rate for unstable samples with satisfactory overall accuracy. Additionally, the proposed method has obvious superiority in computation speed compared with other ensemble models.

**KEY WORDS:** transient stability assessment; machine learning; class imbalance; noise; modified LightGBM

**摘要:** 实际获得的电力系统运行数据, 普遍存在由广域测量系统(WAMS)测量误差造成的噪声, 并且具备类别不平衡的特点, 导致基于机器学习的电力系统暂态稳定评估算法的分类性能受到很大的影响。提出了一种基于改进轻梯度提升机(modified light gradient boosting machine)模型的暂态稳定评估方法, 采用直方图算法对数据进行离散化, 增强模型对噪声的鲁棒性; 在训练中对失稳样本赋予更高的权重, 平衡样

本数量差异造成的影响; 并在损失函数中引入正则项来控制模型复杂度, 减少过拟合, 从而适应电力系统多样的运行情况。在新英格兰 10 机 39 节点系统和美国南卡罗莱纳州 500 节点实际电网上的仿真结果表明, 与其他机器学习方法相比, 所提方法在噪声干扰下不容易过拟合, 具有更好的鲁棒性; 在保持较高总体评估准确率的同时, 对失稳样本具有更高的识别率; 与其他集成学习模型相比, 所提方法在速度上也具有明显优势。

**关键词:** 暂态稳定评估; 机器学习; 类别不平衡; 噪声; 改进 LightGBM

DOI: 10.13335/j.1000-3673.pst.2019.0085

## 0 引言

电力系统的暂态稳定性可以定义为电力系统在受到短路等严重的扰动后保持同步运行的能力<sup>[1]</sup>。近年来, 为了满足日益增长的负荷需求, 电力系统的运行状态更靠近其稳定极限, 使得暂态失稳成为造成停电事故的重要原因之一<sup>[2]</sup>。因此, 对电力系统暂态稳定性做出快速、准确的判断, 为制定合理的调度控制策略提供依据, 是保障电力系统安全稳定运行的重要环节。

目前, 电力系统暂态稳定评估方法主要有 3 类, 包括时域仿真法、直接法和机器学习方法<sup>[3-9]</sup>。时域仿真法将电力系统中各元件用一组微分代数方程组进行描述, 以稳态潮流作为初值, 通过数值积分方法求解方程组, 得到扰动下的发电机转子摇摆曲线, 根据发电机转子功角差来判断系统的暂态稳定性<sup>[3-4]</sup>。该方法能适应各类元件模型且计算结果可靠, 但由于计算速度慢, 无法满足在线评估要求, 只能用于离线计算<sup>[5]</sup>。直接法是一类基于能量观点进行暂态稳定分析的方法, 主要包

基金项目: 国家重点研发计划资助项目(2017YFB0902900); 湖北省电力公司科技项目。

Project Supported by National Key R&D Program of China (2017YFB0902900); Science and Technology Project of Hubei Electric Power Company.

括基于李雅普诺夫的暂态能量函数法、扩展等面积法(EEAC)等方法<sup>[5-8]</sup>。这类方法有明确的物理意义,但存在计算结果保守、确定能量函数困难等问题。

与上述传统的暂态稳定评估方法相比,机器学习方法不需要建立复杂的数学模型,它往往把暂态稳定评估当作一个二分类问题,通过离线训练方式建立电力系统众多物理量与暂态稳定状态之间的映射关系,然后将训练好的模型用于暂态稳定状态预测。基于机器学习的暂态稳定评估方法具有处理电力系统海量数据的能力,并且兼具评估速度快和评估准确性高的特点,目前已有较多研究成果。包括支持向量机<sup>[10-12]</sup>、人工神经网络<sup>[13-14]</sup>、决策树<sup>[15-16]</sup>以及深度学习<sup>[17-18]</sup>等常用机器学习算法在内的分类模型均已被应用到暂态稳定评估中,并取得了较好效果。此外也有一些将暂态稳定评估视为多分类任务的研究方法。文献[13]和[19]基于临界切除时间和故障切除时间的差值定义了暂态稳定裕度指标,将样本分为从稳定、较稳定到较不稳定和 不稳定等多种类别。文献[20]则采用了 2 类支持向量机,将 2 个分类器预测结果不一致样本划分进灰色地带,从而得到 3 种分类结果。这类方法能得到更细致的评估结果,但对系统暂态稳定状态进行多类划分的依据是否具有合理性,仍需要进一步的研究。

目前,基于机器学习的电力系统暂态稳定评估方法依然存在一些不足。由于模型在离线学习时针对的是特定的网络拓扑结构,当系统的拓扑结构发生变化,会带来 2 个问题。其一,针对网络拓扑结构改变前的训练得到的分类模型很可能不再适用。其二,需要重新训练适应于新网络拓扑的模型。由于实际电网规模十分庞大,重新生成模型训练所需的数据集并完成训练,将付出巨大的时间代价。传统的机器学习方法往往缺乏继承性,不过近年来,迁移学习<sup>[21]</sup>理论的发展,为解决这一问题提供了思路。它可以将在一个场景中学习到的知识迁移到另一个场景中应用,意味着可以使基于原始拓扑结构训练得到的模型在拓扑结构改变后的系统中具有良好的泛化性能。此外,如果能找到电力系统中某些拓扑不变量<sup>[22]</sup>来构建暂态稳定评估的输入特征,拓扑结构的变化就不会影响模型的适用性。对于生成数据集会花费大量时间的问题,一方面,如果模型的泛化问题得到解决,就无需再反复生成不同的数据集。另一方面,云计算以及并行计算等技术在电力系统时域仿真,特别是暂态稳定计算中的应用,能有效降低时间成本<sup>[23-24]</sup>。同时,生成对抗神

经网络的提出,也为生产大规模数据集提供了更加便捷的途径<sup>[25]</sup>,在一定程度上减少了机器学习方法实用化的阻碍。

由于电力系统长期处于稳定运行状态,可用于模型训练的不稳定样本数量远远少于稳定样本数量,使得模型对失稳样本特征学习不够充分,导致失稳样本的评估准确率低于稳定样本。此外,通过广域量测系统获得的实际电力系统数据往往含有一定水平的噪声,这会给模型带来过拟合的可能,从而使模型泛化能力下降。现有文献对上述问题的研究不够充分。文献[10]没有计及样本不平衡因素,只给出了总体评估准确率;文献[11]则没有考虑噪声会对失稳样本学习带来影响。文献[26]提出了一种代价敏感学习方式来解决样本不平衡问题,提高了对失稳样本的评估准确性,但算例样本数量较小,说服力有限。文献[27]通过改进 Bootstrap 提高了对失稳样本的识别率,但同样未考虑噪声的影响。文献[28]采用变分自动编码器对数据进行特征抽取,对噪声鲁棒性较好,获得了较高的总体评估准确率,但缺乏噪声条件下对失稳样本评估准确率的分析。实际的电力系统数据中,样本类别不平衡问题和噪声往往同时存在,噪声会带来过拟合,并使模型对少类样本特征的学习受到进一步影响。因此单独考虑其中一个问题不够全面,将导致所构建的评估模型缺乏实际应用价值。

为了弥补以上不足,本文提出了一种基于改进 LightGBM 的暂态稳定评估方法:以采用 Leaf-wise 生长策略的决策树作为基分类器构建高效的集成学习框架,引入直方图算法对原始数据进行离散化以增强对噪声的鲁棒性;考虑到失稳样本数量少但其误分类代价严重,在 LightGBM 的基础上进行改进,通过对失稳样本赋予更高的权重,来改善样本类别不平衡造成的失稳样本识别率低的问题;同时,在原损失函数中加入正则项来限制训练过程中参数的变化程度,从而减少过拟合,以适应电力系统复杂的运行情况。因此,对于含有噪声的电力系统暂态稳定类别不平衡数据,改进后的 LightGBM 具有更好的泛化性能。

## 1 改进 LightGBM 模型

梯度提升决策树<sup>[29]</sup>(gradient boosting decision tree, GBDT)是一个基于决策树<sup>[30]</sup>的集成学习框架,在众多领域表现出了良好的分类正确率,但存在着容易过拟合、训练速度慢等问题。针对这些缺点,Guolin Ke 等人做出了相应的改进,于 2017 年提出

了 GBDT 的一种高效实现方式, 即 LightGBM<sup>[31]</sup>。它采用的主要改进措施包括直方图算法和带深度限制的 Leaf-wise 决策树生长策略, 从而增强了对噪声的鲁棒性, 同时保证了良好的评估准确性和训练速度。为使模型对暂态稳定评估具备更好的适应性, 本文对模型做出了进一步改进。

### 1.1 决策树

决策树是一类常用的机器学习算法。一棵决策树的基本结构如图 1 所示。

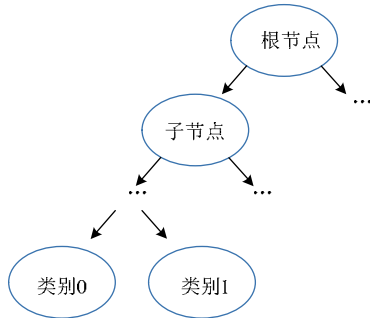


图 1 决策树基本结构

Fig. 1 Basic structure of a decision tree

通常, 一棵决策树包含一个根节点、若干个子节点和若干个叶子节点, 叶子节点对应最终的分类结果, 其他子节点则对应不同的分类特征。根节点包含全部的训练样本, 每个子节点包含根据不同特征进行划分后的样本。决策树学习中最关键的步骤, 就是如何进行特征划分<sup>[32]</sup>。设第  $k$  类样本在训练集  $D$  中的占比为  $p_k (k = 1, 2, \dots, n)$ , 定义信息熵为

$$\text{Ent}(D) = -\sum_{k=1}^n p_k \log_2 p_k \quad (1)$$

若离散特征  $u$  存在  $M$  个可能的取值  $\{u_1, u_2, \dots, u_M\}$ , 则训练集在通过特征  $u$  进行划分后就会产生  $M$  个分支节点。设训练集  $D$  中所有特征  $u$  的取值为  $u_m$  的样本均包含在第  $m$  个分支节点中, 定义信息增益为

$$\text{Gain}(D, u) = \text{Ent}(D) - \sum_{m=1}^M \frac{D^m}{D} \text{Ent}(D_m) \quad (2)$$

特征划分过程主要的依据, 就是使得划分后的式(2)所表示信息增益最大<sup>[30-32]</sup>。

### 1.2 梯度提升决策树

GBDT 是一种将决策树作为基分类器的集成学习框架, 其基本原理描述如下。

给定有限的数据集  $\{y_i, x_i\}_1^N$ , 其中  $x = \{x_1, \dots, x_n\}$  为输入,  $y$  为输出, 定义估计函数  $F(x)$  表示  $x$  到  $y$  的映射关系。构造损失函数  $L(y, F(x))$ , 训练目标即为: 对于所有的  $\{y, x\}$  输入-输出关系构成的联合分布, 找到使得损失函数  $L(y, F(x))$  期

望值最小的函数  $F^*(x)$

$$F^*(x) = \arg \min_F E_{y,x} L(y, F(x)) =$$

$$\arg \min_F E_x [E_y (L(y, F(x))) | x] \quad (3)$$

作为一种集成学习框架, GBDT 的基本思想是由  $M$  个分类准确率较低的弱分类器线性组合构建出一个强分类器, 即

$$F_M(x) = \sum_{m=0}^M \beta_m h(x; a_m) \quad (4)$$

GBDT 所采用的弱分类器  $h(x; a_m)$  是决策树模型。而梯度提升过程则是在训练中使用分段贪婪算法, 以迭代的方式优化下一棵决策树的参数。设  $F(x)$  的初值为  $F_0(x)$ , 其第  $m$  次迭代值为

$$F_m(x) = F_{m-1}(x) + \beta_m h(x; a_m) \quad (5)$$

令  $-g_m(x_i)$  表示第  $m$  次迭代的负梯度方向,  $i = 1, \dots, N$ , 则有

$$-g_m(x_i) = -\left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad (6)$$

$$(\beta_m, a_m) = \arg \min_{\beta, a} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \beta h(x_i; a)) \quad (7)$$

根据式(6)(7), 可得  $a_m$  和  $\beta_m$

$$a_m = \arg \min_{a, \beta} \sum_{i=1}^N [-g_m(x_i) - \beta h(x_i; a)]^2 \quad (8)$$

$$\beta_m = \arg \min_{\beta} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \beta h(x_i; a_m)) \quad (9)$$

如图 2 所示, 通过  $M$  棵决策树的迭代、组合, 即可得到最终的分类器。

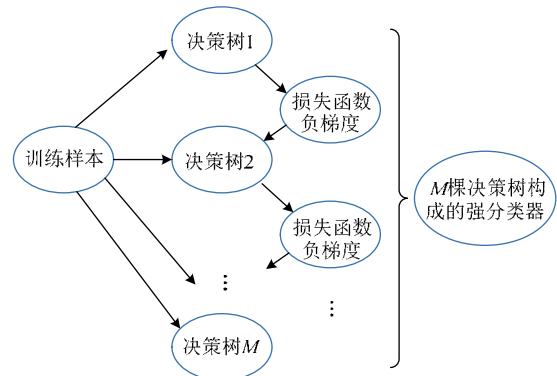


图 2 GBDT 框架

Fig. 2 Framework of GBDT

### 1.3 直方图算法

直方图算法是一种数据离散化处理方法, 它的基本思想是: 对每一维特征, 将连续的浮点型数据划分进离散的  $k$  个范围内, 得到  $k$  个“桶”(bin), 同时构造一个宽度为  $k$  的直方图, 如图 3 所示。

遍历数据时, 把离散后的值作为索引在直方图的每个“桶”中进行累积。遍历一次数据后, 就能

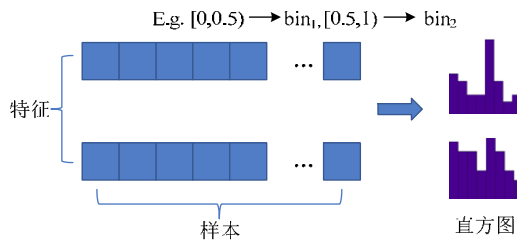


图 3 直方图算法  
Fig. 3 Histogram algorithm

够得到每个“桶”内的累积量。这样一来，在寻找决策树最优分割点时无需逐个对比特征的值，从而大大减少运算量，提高训练速度。

含有噪声的数据会在真实值附近的小范围内波动，当模型逼近能力很强以至于学习到这些波动时，就会造成过拟合的问题。应用直方图算法，对于每一维特征，处于一定区间内的值都会被划分进同一个“桶”，从而获得同样的索引，因此可以显著改善噪声引起的过拟合现象。

1.4 带深度限制的 Leaf-wise 生长策略

决策树是 LightGBM 所使用的基分类器，而大部分决策树采用的是 Level-wise 生长策略，即按层生长，如图 4 所示。事实上，很多节点分裂后所获得的信息增益很低，对分类结果影响较小，全部分裂反而会带来很大的计算代价，降低模型训练速度。因此 LightGBM 采用了一种带深度限制的 Leaf-wise 策略，即按叶生长，如图 5 所示。

按叶生长，就是在每一次分裂节点时，选择增益最大的叶子节点进行生长。在分裂次数相同的情况下，与 Level-wise 方式相比，采用 Leaf-wise 可以得到更低的误差。LightGBM 通过额外的参数树的最大深度(max\_depth)和叶子数(num\_leaves)来限制树的深度，从而控制了模型的复杂度，避免出现

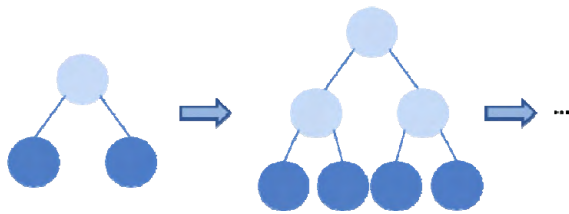


图 4 Level-wise 生长方式  
Fig. 4 Level-wise growth

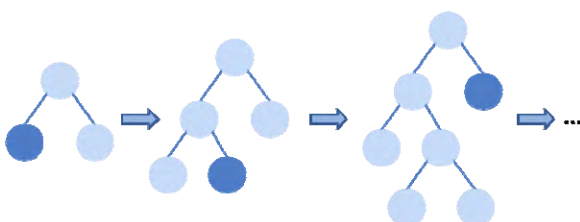


图 5 带深度限制的 Leaf-wise 生长方式  
Fig. 5 Leaf-wise growth with depth constrain

过拟合。在提升训练速度的同时，保证了模型的泛化能力。

1.5 LightGBM 改进方法

对于暂态稳定评估而言，将不稳定样本错分类成稳定样本与将稳定样本错分类成不稳定样本，造成的后果有很大差异<sup>[26]</sup>，前者的代价明显远高于后者。由于实际电力系统中的失稳运行样本数据稀少，常规分类器会在训练过程中更多地学习到稳定样本的特征，导致对失稳样本的识别能力不足。同时，电力系统复杂的运行情况会增加样本的多样性，从而导致模型出现过拟合的问题。为了解决上述问题，对 LightGBM 的损失函数进行了改进。首先在损失函数中引入 L<sub>2</sub> 正则项<sup>[32-33]</sup>来减少过拟合，然后对失稳样本赋予更高的权重来处理样本不平衡问题，具体改进方法如下。

对第 m 棵树，原损失函数为

$$L_{original} = -\frac{1}{N} \sum_{i=1}^N L(y_i, F_{m-1}(x_i; A_{m-1})) \quad (10)$$

其中  $F_{m-1}(x_i; A_{m-1})$  表示由前 m-1 棵树组成的模型在参数为  $A_{m-1}$  的条件下对输入  $x_i$  的预测值，这里的  $A_{m-1}$  包括了前 m-1 棵树的参数  $a_1, \dots, a_{m-1}$ 。 $L(y_i, F_{m-1}(x_i; A_{m-1}))$  是描述真实值  $y_i$  与当前模型预测值误差的函数，对 LightGBM 而言是对数损失函数。

改进后第 m 棵树的损失函数为

$$L_{modified} = -\frac{1}{N} (\sum_{i=1}^N \alpha_i L(y_i, F_{m-1}(x_i; A_{m-1})) + \frac{\lambda}{2} \|a\|_2^2) \quad (11)$$

其中  $\lambda$  是正则化系数，除以 2 是为了与平方项求导后相消，简化计算。系数  $\alpha_i$  为：

$$\alpha_i = \begin{cases} c, & y_i = 1 \\ 1, & y_i = 0 \end{cases} \quad (12)$$

其中， $y_i = 0$  表示稳定样本， $y_i = 1$  表示失稳样本。 $c$  是一个大于 1 的常数，具体取值与样本类别比例有关。

对失稳样本加权，可以使得得到更大权重的失稳样本在训练时获得更大的梯度、更受重视，从而平衡样本数量差异带来的影响，提高模型对失稳样本的评估准确性。本文采用的 L<sub>2</sub> 正则化又被称作岭回归<sup>[34]</sup>，常用作多种模型防止过拟合的手段。根据机器学习领域的基本原则“奥卡姆剃刀定律”<sup>[32]</sup>，设置正则项之所以能减少过拟合，是因为它限制了迭代训练过程中模型的变化程度，起到了控制模型复杂度的作用，使得训练后的模型有更好的平滑性。由于模型对电力系统高维数据的拟合难以用可视化的方式展现，本文以平面上的多项式

曲线对数据点的拟合来解释引入正则项减少过拟合的原理，如图6—7所示。本质上，曲线的拟合方式就相当于模型对训练集的学习结果。图6中的多项式曲线波动程度较大，平滑性较差，与图7中的曲线相比需要更大的拟合系数。虽然图6所示的拟合方式对原数据点逼近程度更高，但一个与原数据独立同分布的未知数据显然更有可能落在图7所示的曲线上。相应地推广到高维空间中，参数复杂度更低的模型在面对未知样本时往往具有更好的泛化能力。

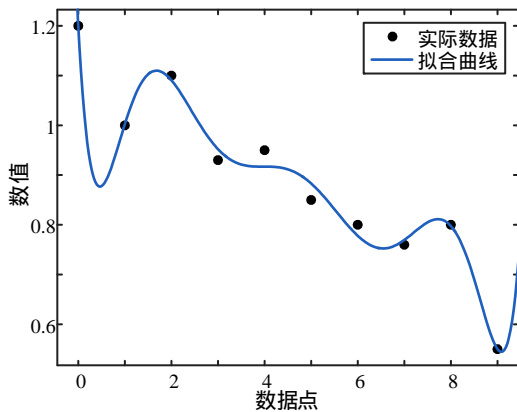


图6 参数较大的拟合情况

Fig. 6 Fitting with larger parameters

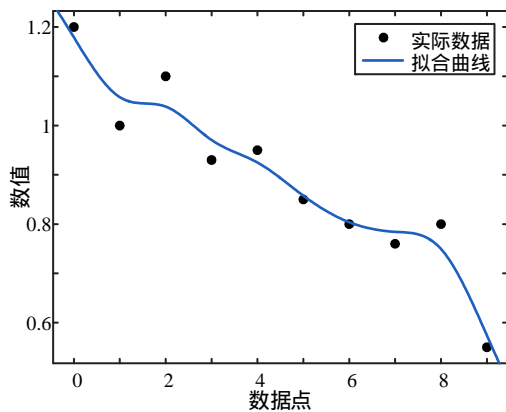


图7 参数较小的拟合方式

Fig. 7 Fitting with smaller parameters

## 2 暂态稳定评估模型

### 2.1 输入特征构建

基于机器学习的暂态稳定评估，本质就是建立输入特征与暂态稳定状态的映射关系，因此输入特征的选择十分重要。故障切除后，系统不会再受到外界干扰，即系统的稳定状态将完全由系统内部各元件的物理特性决定，因此故障切除瞬间各物理量的值就可以完全反映系统未来的运行状态。而故障后实际的暂态过程中，难以精确测量发电机q轴电势这类状态量<sup>[35]</sup>。因此，本文选择在故障切除瞬间能够准确获得的若干物理量构建输入特征集，各特

征类型如表1所示。其中，各类特征包含的具体特征维数不是固定值，而与实际电网规模有关。

表1 输入特征类型	
Tab. 1 Categories of input features	
序号	特征类型
1	故障切除瞬间各发电机有功功率
2	故障切除瞬间各发电机无功功率
3	故障切除瞬间各母线电压幅值
4	故障切除瞬间各母线电压相角
5	故障切除瞬间各线路有功功率
6	故障切除瞬间各线路无功功率
7	故障切除瞬间各负荷有功功率
8	故障切除瞬间各负荷无功功率

### 2.2 暂态稳定评估流程

基于LightGBM的暂态稳定评估与其他机器学习方法<sup>[10-15]</sup>类似，主要由离线训练和在线评估两部分构成。离线训练包括数据集生成、分类模型训练和模型性能评估3个环节，而在线评估则包含在线获取特征数据、将特征数据输入训练好的模型和预测系统暂态稳定状态等步骤。完整的电力系统暂态稳定评估流程如图8所示。

### 2.3 模型性能评价

为了衡量训练后模型的好坏，需要对模型性能进行评估。由于样本中稳定样本数量远多于失稳样本数量，只要模型能够对稳定样本有较好的分类准确率，即使对失稳样本识别率很低，也可以得到很

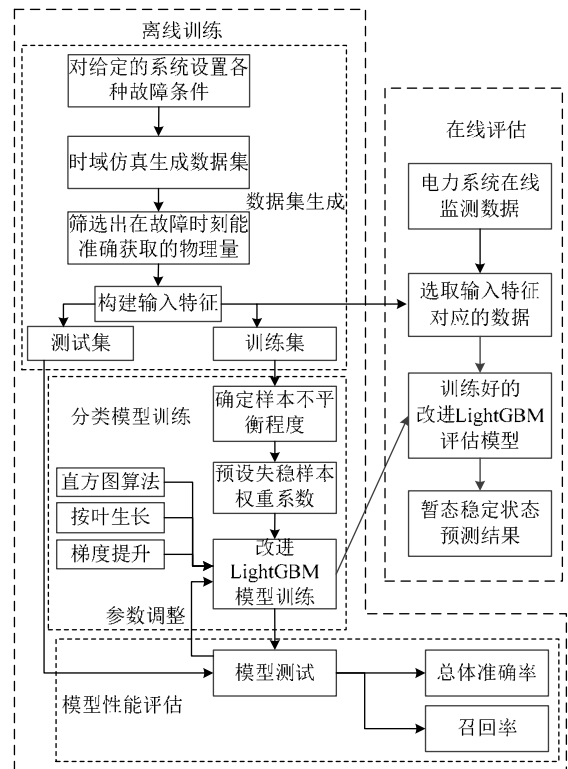


图8 电力系统暂态稳定评估流程

Fig. 8 Process of power system transient stability assessment

高的总体准确率。但在实际的电力系统调度中,对失稳情况的判断反而关注程度更高。因此本文在评估模型性能时,除了传统的总体准确率指标,也同时考虑反映对失稳样本识别情况的召回率指标。

定义总体准确率(accuracy,  $A_{cc}$ )指标和召回率(recall,  $R_{cc}$ )指标<sup>[32]</sup>

$$A_{cc} = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \quad (13)$$

$$R_{cc} = \frac{T_P}{T_P + F_N} \quad (14)$$

其中:  $T_P$  表示被正确分类的不稳定样本数量;  $T_N$  表示被正确分类的稳定样本数量;  $F_N$  表示被错误分类的不稳定样本数量;  $F_P$  表示被错误分类的稳定样本数量。由定义可知,总体准确率表示的是所有样本中被正确分类的比例,召回率表示不稳定样本中被正确分类的比例。

### 3 算例分析

#### 3.1 新英格兰 39 节点系统

##### 3.1.1 样本生成与参数确定

为了验证所提方法的有效性,本文首先选择新英格兰 10 机 39 节点系统作为测试系统<sup>[36]</sup>。该系统包含 10 台发电机、39 条母线和 46 条线路。采用 PSS/E<sup>[37]</sup>软件进行仿真,发电机设置为 GENROU 模型,负荷采用恒阻抗模型,仿真时间步长为 0.0083s。以 5% 为步长,共设置 70%~120% 范围内的 11 种负荷水平,发电机出力随负荷水平进行相应调整。在各母线上及线路的 20%、40%、60%、80% 处设置三相接地短路故障,故障持续时间分别设置为 0.1、0.2 或 0.3 s。仿真总时长为 20 s,若仿真结束时任意两发电机间转子功角差大于 360°,则认为样本暂态失稳<sup>[38]</sup>。为了模拟样本不平衡的情况,进行仿真后随机剔除部分失稳样本。共得到 4962 个样本,其中稳定样本 3890 个,失稳样本 1072 个。稳定样本与失稳样本的比例约 3.6:1。为了保证模型的泛化性能,随机选择样本中的 70% 作为训练集,剩下 30% 作为测试集。

根据 IEEE C37.118 标准,PMU 相量测量误差应小于 1%<sup>[39]</sup>。按照信噪比的定义,信噪比为 20dB 时即意味着数据中含有 1% 的噪声。为了验证所提模型在不同噪声条件下的泛化能力,分别向原始数据中添加信噪比为 40、30、20 和 15dB 的高斯白噪声来模拟电力系统广域测量系统误差,各信噪比对应的测量误差分别为 0.01%、0.1%、1%、和约 3%。

影响改进 LightGBM 模型评估性能的参数主要

有基分类器数目、叶子数以及失稳样本权重  $c$  和正则化系数  $\lambda$ 。本文在考虑数据噪声含量最高(即信噪比为 15dB)的条件下,采用网格搜索法<sup>[40]</sup>确定以上参数。最终得到的最佳参数组合是:基分类器数量为 250,叶子数为 4,失稳样本权重为 5,正则化系数为 0.2。

##### 3.1.2 仿真结果分析

为进一步证明本文所提方法的优越性,将改进 LightGBM(mLightGBM)与原 LightGBM 模型以及其他常用机器学习模型的评估结果进行对比。测试模型均基于开源机器学习库 Scikit-learn<sup>[41]</sup>搭建。参数使用 Scikit-learn 的默认值。测试情况如表 2 和表 3 所示。

表 2 新英格兰 39 节点系统评估总体准确率  
Tab. 2 Assessment accuracy on New England 39-bus system

模型	不同噪声水平下的总体准确率/%				
	无噪声	40dB	30dB	20dB	15dB
DT	96.91	96.59	96.64	95.73	94.07
KNN	98.36	98.36	98.31	98.29	98.03
ANN	98.30	98.52	98.37	97.80	97.01
SVM	99.34	99.07	98.77	98.04	97.00
RF	99.07	98.96	98.68	98.58	97.92
GBDT	99.15	98.95	98.88	98.49	98.29
LightGBM	99.53	99.32	99.21	99.05	98.85
mLightGBM	99.53	99.32	99.20	99.04	98.91

表 3 新英格兰 39 节点系统评估召回率  
Tab. 3 Assessment recall on New England 39-bus system

模型	不同噪声水平下的召回率/%				
	无噪声	40dB	30dB	20dB	15dB
DT	95.61	95.73	95.75	93.88	92.32
KNN	95.98	95.75	95.59	95.15	93.17
ANN	96.77	95.50	94.43	94.02	93.37
SVM	97.70	97.43	97.24	95.80	93.56
RF	97.92	97.30	96.61	95.69	93.90
GBDT	97.99	97.70	97.41	96.55	95.56
LightGBM	99.36	98.56	98.09	97.46	96.20
mLightGBM	99.36	98.73	98.49	98.03	97.85

表 2 和表 3 中,前 4 种模型是个体学习器,其余为集成学习模型。总体来看,由于集成策略在训练过程中有效降低了方差<sup>[32]</sup>,集成学习模型相对拥有更好的泛化性能。而 DT、KNN 和 ANN 等简单模型由于结构简单,学习能力有限,且没有考虑样本平衡带来的影响,总体准确率和召回率都处于较低水平。SVM 在数据不含噪声的情况下,可以得到良好的总体准确率和召回率。但随着噪声的增大,其评估性能迅速降低,说明 SVM 对噪声鲁棒性较差。同为集成模型的 RF、GBDT 和 LightGBM,在数据噪声含量较小时总体准确率相近。但由于设置了失稳样本权重,改进 LightGBM 具有最高的召回

率。当信噪比不断增大时, RF 和 GBDT 都受到了较大的影响, 而改进 LightGBM 则表现出了较强的鲁棒性。在信噪比为 15dB 的极端条件下, 改进 LightGBM 的总体准确率比 RF 和 GBDT 分别高 0.91% 和 0.54%, 召回率比 RF 和 GBDT 分别高 3.8% 和 2.14%。在噪声逐渐增大的过程中, mLightGBM 的召回率相比标准 LightGBM 的提升幅度也越大, 从而证明了改进的必要性和有效性。

为了进一步说明不平衡样本对各种分类模型的影响, 通过随机剔除部分样本, 设置不同的稳定样本与失稳样本比例, 得到相应的测试结果如图 9 所示。由于样本的不平衡程度主要影响模型对失稳样本的学习和识别能力, 此时应主要关注召回率指标的变化情况。总体来说, 样本不平衡程度越严重, 各类模型的召回率下降得越多。这是因为稳定样本多于失稳样本时, 模型在训练过程中为了最小化总体训练误差, 倾向于将一个样本判断为稳定样本。而改进 LightGBM 根据不同的类别比例, 在损失函数中对失稳样本赋予不同权重, 较好地平衡了分类边界, 因此始终保持了最高的召回率。可以发现, 不平衡程度比较严重时, 改进 LightGBM 具有更显著的优越性。

由以上分析可知, 与其他方法相比, 本文所提出的改进 LightGBM 模型能有效改善噪声和样本不平衡问题对暂态稳定评估带来的影响。

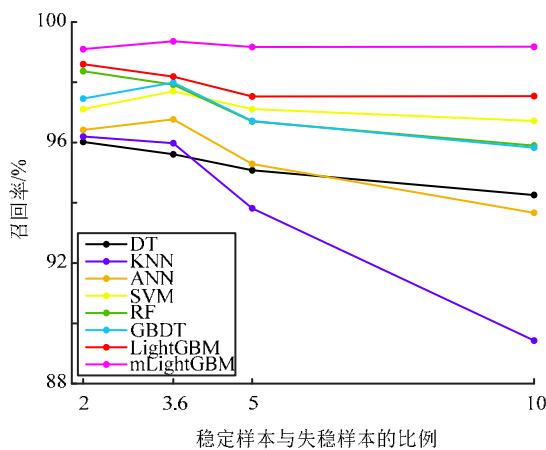


图 9 不平衡样本集对各种模型影响

Fig. 9 Impacts of imbalanced datasets on different models

### 3.2 美国南卡罗莱纳州 500 节点电网

#### 3.2.1 样本生成与参数确定

为了进一步验证所提方法在大电网暂态稳定评估中的有效性, 采用美国南卡罗莱纳州 500 节点实际电网进行测试<sup>[38]</sup>。美国南卡罗莱纳州 500 节点电网包含 56 台发电机和 500 条母线。采用 PSSE/E 软件进行仿真, 以 10% 为步长, 共设置 70%~110%

范围内的 5 种负荷水平, 发电机出力随负荷水平进行相应调整。在各母线上及线路 20%、40%、60%、80% 处设置三相接地短路故障, 故障持续时间分别设置为 0.1、0.3 或 0.5s。为了模拟更严重的样本不平衡情况, 进行仿真后随机剔除部分失稳样本。共得到 9838 个样本, 其中稳定样本 8967 个, 失稳样本 871 个。稳定样本与失稳样本的比例约为 10:1。随机选择样本中的 70% 作为训练集, 剩下 30% 作为测试集。

参数确定过程与 39 节点系统相同。得到的最佳参数组合: 基分类器数目 250, 叶子数为 4, 失稳样本权重为 10, 正则化系数为 0.1。

设置失稳样本权重是为了平衡 2 类样本数量差异对模型学习及预测结果的影响, 而不同的权重系数会产生不同的效果。当数据中噪声含量为 1% (即 20dB) 时, 其他参数不变, 模型召回率和总体准确率随失稳样本权重的变化如图 10 所示。权重系数小于 10 时, 召回率随其增大而上升, 并与总体准确率越来越接近。若继续增大权重系数至 15 或 20, 召回率不再提升, 只会导致总体准确率略有下降。因此 10 是比较合理的取值, 与训练集中稳定样本与失稳样本的比例相近。

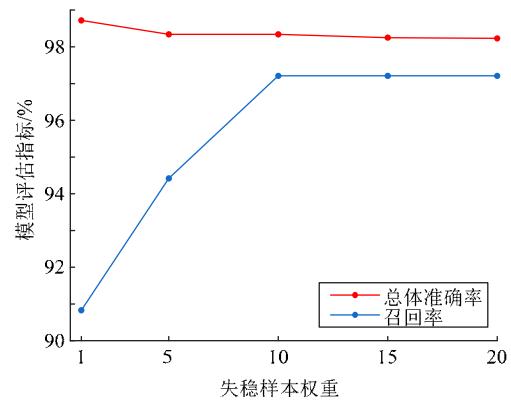


图 10 不同失稳样本权重的影响

Fig. 10 Effects of different unstable sample weights

#### 3.2.2 仿真结果分析

由于 500 节点电网的规模远高于 39 节点电网, 输入特征的维度也相应增加。对泛化性能不佳的模型而言, 更容易出现过拟合。为了验证模型在大数据样本中的有效性, 在南卡罗莱纳州 500 节点电网上进行了测试。模型参数使用 Scikit-learn 的默认值, 结果如图 11 所示。

由图 11 可知, 与新英格兰 39 节点系统上的测试结果相比, 更大的样本总量使得各模型学习更加充分, 因此能得到较高的总体准确率。但由于样本类别不平衡问题更加突出, 除改进 LightGBM 模型

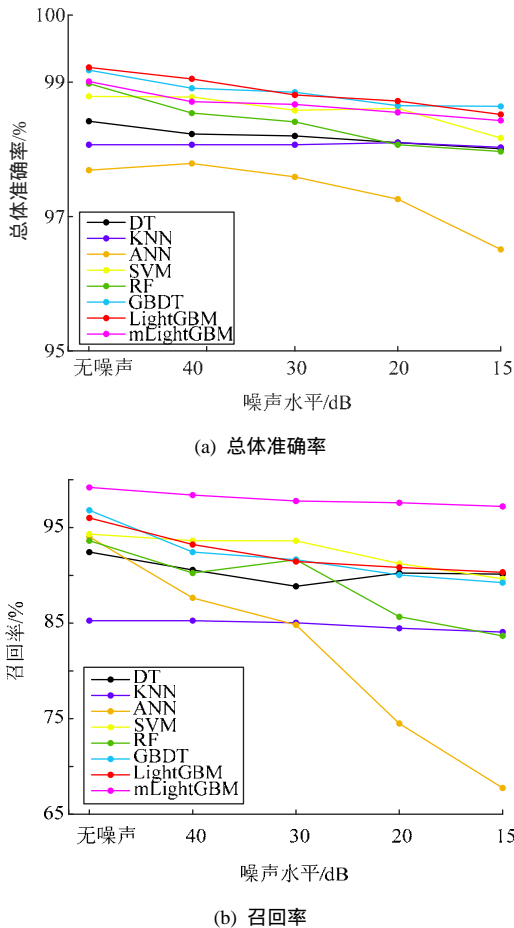


图 11 南卡罗莱纳州 500 节点电网评估结果  
Fig. 11 Assessment results on 500-bus grid in State of South Carolina

外的各模型召回率均有明显下降。尤其是随着噪声的增大，除改进 LightGBM 外的各模型召回率逐渐降至 90% 左右或更低，说明噪声进一步影响了这些模型对少类样本的学习。这是因为模型需要以更大的参数以及更不平滑的方式去拟合含有噪声的少类样本。通过 1.5 节的分析可知，更大的参数更容易造成过拟合，导致模型泛化性能下降。而仿真结果表明，采取了多项改进措施的 mLightGBM 对以上问题有良好的鲁棒性。

类别严重不平衡的含噪声高维数据更符合真实的电力系统运行数据，因此 SVM 等传统模型在实际的大电网暂态稳定评估中难以得到应用。改进 LightGBM 模型在这种条件下仍然能保持较高的总体准确率和召回率，因此具有更高的实际应用价值。

### 3.3 时间复杂度

集成模型的训练时间过长是阻碍其广泛应用的一个重要原因。LightGBM 由于采用了多种加速措施，使其获得了远超其他集成方法的训练速度，这对处理电力系统海量数据具有重要意义。为了对比各种方法的训练速度，将各种集成模型的基分类

器数目均设置为 250，SVM 和 DT 使用 scikit-learn 默认参数，在美国南卡罗莱纳州 500 节点电网的样本集上进行测试，结果如表 4 所示。

由表 4 可见，mLightGBM 和 LightGBM 速度相近，均大幅缩短了集成学习的训练时间，甚至比作为个体分类器的 SVM 和 DT 效率更高。当电网运行方式发生改变，mLightGBM 能更快地完成对模型参数的调整，从而拥有更好的适应性。

模型	训练时间/s
RF	211.47
GBDT	658.09
SVM	31.55
DT	53.99
LightGBM	15.77
mLightGBM	15.62

## 4 结论

本文考虑了电力系统运行数据含有噪声和样本不平衡的问题，提出了一种基于改进 LightGBM 的电力系统暂态稳定评估方法，在新英格兰 39 节点系统和南卡罗莱纳州 500 节点电网上进行了仿真分析。结果表明，在样本类别不平衡的情况下，噪声会严重影响传统分类模型对失稳样本的评估准确率，从而使其失去实用意义。与其他基于机器学习的评估方法相比，改进 LightGBM 在噪声和样本类别不平衡条件下，有效改善了过拟合问题，保证了较高的总体评估准确率，对失稳样本识别能力更好，并且具有更低的时间复杂度，尤其在处理高维数据时优势明显，因此在电力系统中拥有更显著的应用价值。

当电力系统的运行方式和拓扑结构发生变化，之前训练得到的模型可能不再可用，因此如何利用迁移学习等技术使模型具有继承性和更广泛的适应性，是未来研究的重点。同时，如何构造不变特征用于模型学习，也是一个值得探究的方向。

## 参考文献

- [1] IEEE/CIGRE joint task force on stability terms and definitions. Definition and classification of power system stability [J]. IEEE Transactions on Power Systems, 2004, 19(2): 1387-1401.
- [2] Yu J, Hill D, Lam A, et al. Intelligent time-adaptive transient stability assessment system[J]. IEEE Transactions on Power Systems, 2017, 33(1): 1049-1058.
- [3] 常辉, 刘文颖, 行舟, 等. 电力系统暂态稳定计算的在线应用[J]. 电网技术, 2007, 31(13): 54-58, 74. Chang Hui, Liu Wenying, Xing Zhou, et al. Online application of power system transient stability computation[J]. Power System Technology, 2007, 31(13): 54-58, 74(in Chinese).



- [4] Tang C K, Graham C E, El-Kady M, et al. Transient stability index from conventional time domain simulation[J]. IEEE Transactions on Power Systems, 1994, 9(3): 1524-1530.
- [5] 苟竞, 刘俊勇, Gareth Taylor, 等. 基于发电机对暂态势能集的电力系统暂态稳定快速评估[J]. 电网技术, 2015, 39(2): 464-471. Gou Jing, Liu Junyong, Gareth Taylor, et al. Fast assessment of power system transient stability based on transient potential energy set of pairwise generators[J]. Power System Technology, 2015, 39(2): 464-471(in Chinese).
- [6] 罗恒, 刘涤尘, 史秋芸, 等. 基于暂态能量函数的核电机组接入电网暂态稳定性研究[J]. 电网技术, 2013, 37(1): 119-125. Luo Heng, Liu Dichen, Shi Qiuyun, et al. Study on transient stability of nuclear power plant connected to power grid based on transient energy function[J]. Power System Technology, 2013, 37(1): 119-125(in Chinese).
- [7] 薛禹胜. EEAC与直接法的机理比较[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[J]. Automation of Electric Power Systems, 2001, 25(11): 6-11(in Chinese).
- [8] Xue Y, Van Cutsem T, Ribbens-Pavella M. Extended equal area criterion justifications, generalizations, applications[J]. IEEE Transactions on Power Systems, 1989, 4(1): 44-52.
- [9] 吴为, 汤涌, 孙华东, 等. 基于广域量测信息的电力系统暂态稳定研究综述[J]. 电网技术, 2012, 36(9): 81-87. Wu Wei, Tang Yong, Sun Huadong, et al. A survey on research of power system transient stability based on wide-area measurement information[J]. Power System Technology, 2012, 36(9): 81-87(in Chinese).
- [10] 陈厚合, 王长江, 姜涛, 等. 基于投影能量函数和 Pin-SVM 的电力系统暂态稳定评估[J]. 电工技术学报, 2017, 32(11): 67-76. Chen Houhe, Wang Changjiang, Jiang Tao, et al. Transient stability assessment in bulk power grid using projection energy function and support vector machine with pinball loss[J]. Transactions of Chinese Electrotechnical Society, 2017, 32(11): 67-76(in Chinese).
- [11] 戴远航, 陈磊, 张玮灵, 等. 基于多支持向量机综合的电力系统暂态稳定评估[J]. 中国电机工程学报, 2016, 36(5): 1173-1180. Dai Yuanhang, Chen Lei, Zhang Weiling, et al. Power system transient stability assessment based on multi-support vector machines[J]. Proceedings of the CSEE, 2016, 36(5): 1173-1180(in Chinese).
- [12] Wang B, Fang B, Wang Y, et al. Power system transient stability assessment based on big data and the core vector machine[J]. IEEE Transactions on Smart Grid, 2016, 7(5): 2561-2570.
- [13] 姚德全, 贾宏杰, 赵帅. 基于复合神经网络的电力系统暂态稳定评估和裕度预测[J]. 电力系统自动化, 2013, 37(20): 41-46. Yao Dequan, Jia Hongjie, Zhao Shuai. Power system transient stability assessment and stability margin prediction based on compound neural network[J]. Automation of Electric Power Systems, 2013, 37(20): 41-46(in Chinese).
- [14] 汤必强, 邓长虹, 刘丽芳. 复合神经网络在电力系统暂态稳定评估中的应用[J]. 电网技术, 2004, 28(15): 62-66. Tang Biqiang, Deng Changhong, Liu Lifang. Application of compound neural network in power system transient stability assessment[J]. Power System Technology, 2004, 28(15): 62-66(in Chinese).
- [15] 王康, 孙宏斌, 张伯明, 等. 基于二维组合属性决策树的暂态稳定评估[J]. 中国电机工程学报, 2009, 29(S1): 17-24. Wang Kang, Sun Hongbing, Zhang Boming, et al. Transient stability assessment based on 2D combined attribute decision tree[J]. Proceedings of the CSEE, 2009, 29(S1): 17-24(in Chinese).
- [16] 叶圣永, 王晓茹, 刘志刚, 等. 电力系统暂态稳定评估组合模型比较[J]. 电网技术, 2008, 32(23): 19-23. Ye Shengyong, Wang Xiaoru, Liu Zhigang, et al. Comparison research on combined models for power system transient stability assessment[J]. Power System Technology, 2008, 32(23): 19-23(in Chinese).
- [17] 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).
- [18] 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).
- [19] 尹雪燕, 闫炯程, 刘玉田, 等. 基于深度学习的暂态稳定评估与严重程度分级[J]. 电力自动化设备, 2018, 38(5): 64-69. Yin Xueyan, Yan Jiongcheng, Liu Yutian, et al. Deep learning based transient stability assessment and severity grading[J]. Electric Power Automatic Equipment, 2018, 38(5): 64-69(in Chinese).
- [20] 张玮灵, 胡伟, 闵勇, 等. 稳定域概念下考虑保守性的电力系统在线暂态稳定评估方法[J]. 电网技术, 2016, 40(4): 992-998. Zhang Weiling, Hu Wei, Min Yong, et al. Conservative online transient stability assessment in power system based on concept of stability region[J]. Power System Technology, 2016, 40(4): 992-998(in Chinese).
- [21] 杨挺, 赵黎媛, 王成山. 人工智能在电力系统及综合能源系统中的应用综述[J]. 电力系统自动化, 2019, 43(1): 2-14. Yang Ting, Zhao Liyuan, Wang Chengshan. Review on application of artificial intelligence in power system and integrated energy system[J]. Automation of Electric Power Systems, 2019, 43(1): 2-14(in Chinese).
- [22] 李天云, 李善颖. 拓扑不变量在电力系统混沌特征分析中的应用[J]. 中国电机工程学报, 2002, 22(10): 51-56. Li Tianyun, Li Shanying. The application of topological invariants in chaos characteristics' analysis of power system[J]. Proceedings of the CSEE, 2002, 22(10): 51-56(in Chinese).
- [23] 董毅峰, 王彦良, 韩佶, 等. 电力系统高效电磁暂态仿真技术综述[J]. 中国电机工程学报, 2018, 38(8): 2213-2231. Dong Yifeng, Wang Yanliang, Han Ji, et al. Review of high efficiency digital electromagnetic transient simulation technology in power system[J]. Proceedings of the CSEE, 2018, 38(8): 2213-2231 (in Chinese).
- [24] 江涵. 大规模电力系统暂态稳定并行计算研究[D]. 杭州: 浙江大学, 2012.
- [25] 陈文兵, 管正雄, 陈允杰. 基于条件生成式对抗网络的数据增强方法[J]. 计算机应用, 2018, 38(11): 3305-3311. Chen Wenbing, Guan Zhengxiong, Chen Yunjie. Data augmentation method based on conditional generative adversarial net model[J]. Journal of Computer Applications, 2018, 38(11): 3305-3311(in Chinese).
- [26] 陈振, 肖先勇, 李长松, 等. 基于代价敏感极端学习机的电力系统暂态稳定评估方法[J]. 电力自动化设备, 2016, 36(2): 118-123. Chen Zhen, Xiao Xianyong, Li Changsong, et al. Power system transient stability assessment based on cost-sensitive extreme learning machine[J]. Electric Power Automatic Equipment, 2016, 36(2): 118-123(in Chinese).
- [27] 周艳真, 吴俊勇, 于之虹, 等. 用于电力系统暂态稳定预测的支持向量机组合分类器及其可信度评价[J]. 电网技术, 2017, 41(4):

- 1188-1196 .  
Zhou Yanzhen , Wu Junyong , Yu Zhihong , et al . Support vector machine ensemble classifier and its confidence evaluation for transient stability prediction of power systems[J] . Power System Technology , 2017 , 41(4) : 1188-1196(in Chinese) .
- [28] 周悦 , 谭本东 , 李淼 , 等 . 基于深度学习的电力系统暂态稳定评估方法[J] . 电力建设 , 2018 , 39(2) : 103-108 .  
Zhou Yue , Tan Bendong , Li Miao , et al . Transient stability assessment of power system based on deep learning technology[J] . Electric Power Construction , 2018 , 39(2) : 103-108(in Chinese) .
- [29] Friedman J H . Greedy function approximation : a gradient boosting machine[J] . The Annals of Statistics , 2001 , 29(5) : 1189-1232 .
- [30] Quinlan J R . Induction of decision trees[J] . Machine Learning , 1986 , 1(1) : 81-106 .
- [31] Ke G , Meng Q , Finley T , et al . LightGBM : a highly efficient gradient boosting decision tree[C]//Advances in Neural Information Processing Systems , 2017 .
- [32] 周志华 . 机器学习[M] . 北京 : 清华大学出版社 , 2016 .
- [33] Chen T , Guestrin C . XGBoost : a scalable tree boosting system[C]//ACM SIGKDD International Conference on Knowledge Discovery & Data Mining , 2016 .
- [34] Hoerl A , Kennard R . Ridge regression : biased estimation for nonorthogonal problems[J] . Technometrics , 1970 , 12(1) : 55-67 .
- [35] 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) .
- [36] Pai M A . Energy function analysis for power system stability [M] . Boston : Spinger , 1989 .
- [37] 程华 , 徐政 . PSASP 与 PSS/E 稳定计算的模型与结果比较[J] . 电网技术 , 2004 , 28(5) : 1-4 .  
Cheng Hua , Xu Zheng . Comparison of mathematical models for transient stability calculation in PSASP and PSS/E and corresponding calculation results[J] . Power System Technology , 2004 , 28(5) : 1-4(in Chinese) .
- [38] 谭本东 , 杨军 , 赖秋频 , 等 . 基于改进 CGAN 的电力系统暂态稳定样本增强方法[J] . 电力系统自动化 , 2019 , 43(1) : 149-157 .  
Tan Bendong , Yang Jun , Lai Qiupin , et al . Data augment method for power system transient stability assessment based on improved conditional generative adversarial network[J] . Automation of Electric Power Systems , 2019 , 43(1) : 149-157(in Chinese) .
- [39] He M , Vittal V , Zhang J . Online dynamic security assessment with missing PMU measurements : a data mining approach[J] . IEEE Transactions on Power Systems , 2013 , 28(2) : 1969-1977 .
- [40] 纪昌明 , 周婷 , 向腾飞 , 等 . 基于网格搜索和交叉验证的支持向量机在梯级水电系统隐随机调度中的应用[J] . 电力自动化设备 , 2014 , 34(3) : 125-131 .  
Ji Changming , Zhou Ting , Xiang Tengfei , et al . Application of support vector machine based on grid search and cross validation in implicit stochastic dispatch of cascaded hydropower stations[J] . Electric Power Automatic Equipment , 2014 , 34(3) : 125-131(in Chinese) .
- [41] Scikit-learn : [https://scikit-learn.org/stable/\[2018-12-30\]](https://scikit-learn.org/stable/[2018-12-30]) .



周挺

收稿日期：2019-01-13。

作者简介：

周挺(1997)，男，硕士研究生，研究方向为人工智能在电力系统中的应用，E-mail：starry\_zt@whu.edu.cn；

杨军(1977)，男，教授，通信作者，博士生导师，研究方向为电力系统运行与控制等，E-mail：JYang@whu.edu.cn。

(责任编辑 王晔)