

基于Fisher Score特征选择的电力系统暂态稳定评估方法

李 鹏,董鑫剑,孟庆伟,陈继明

(中国石油大学(华东) 新能源学院 电气工程系,山东 青岛 266580)

摘要:针对不同电气输入特征与电力系统暂态稳定关联程度不同以及当输入特征受到干扰时评估准确率明显下降的问题,提出一种基于Fisher Score特征选择的电力系统暂态稳定评估方法。设计一种面向电力系统暂态稳定评估二分类问题的样本特征Fisher Score值计算方案;通过Fisher Score值排序有效区分重要特征与冗余特征、噪声特征与非噪声特征;将选择的电气特征输入不同机器学习模型中进行训练和评估。新英格兰39节点系统和IEEE 145节点系统的仿真结果表明,所提特征选择方案能有效筛选电力系统暂态稳定评估中重要度高的特征,提升了评估模型的预测性能。

关键词:电力系统;暂态稳定评估;特征选择;Fisher Score算法

中图分类号:TM712

文献标志码:A

DOI:10.16081/j.epae.202212002

0 引言

电力系统暂态稳定评估(transient stability assessment, TSA)是提高电力系统安全稳定性的的重要手段。随着我国电力系统电力电子化程度的提高和可再生能源发电的大量接入,电力系统在运行过程中的暂态稳定状态存在更大的不确定性,对暂态稳定的分析调控难度也随之增大^[1]。相较于传统的时域仿真法^[2-3],基于机器学习的TSA方法具有计算效率高、评估速度快的优点。目前,对基于机器学习的TSA方法的研究主要关注电力系统暂态稳定预测模型的设计,而较少关注输入电气特征集的选择和构建,这些仍需凭借人工经验。文献[4]将转子角轨迹簇的时序特征作为TSA的输入特征,在一定程度上解决了特征紧密性差、整体计算效率低的问题;文献[5]从能量函数的角度出发,将功角/电压幅值/电压相角/有功功率/无功功率的原始值、积分量、微分量等时序数据作为原始输入特征,有效提高了量测数据中暂态信息的利用率;文献[6]将“源-网”状态量作为评估模型的输入特征,能够对暂态稳定状态的边界进行有效评估。由于人工选取的电气特征可能存在一定的冗余特征,这些冗余特征对电力系统TSA的价值较小,并且若冗余特征过多,则会导致后续机器预测模型的训练效率降低,此外,在某些电气特征混入噪声干扰之后,若不加选择地将其全部作为后续电力系统TSA模型的输入特征,则会使得模型的最终预测准确率大幅降低^[7],因此,如何针对电力系统TSA问题进行输入电气特征的有效选择是一个亟待深入研究的重要问题。

近年来,在传统的数据挖掘领域,国内外学者对数据特征的选择问题进行了大量的研究^[8-10];文献

[8]利用极限梯度提升算法对特征进行降维,再利用实体嵌入网络对其进行分类,提高了模型的评估性能;文献[9]利用堆叠变分自动编码器,通过设定各发电机最大功角差的阈值对特征进行提取,大幅降低了模型的误判率;文献[10]提出一种聚类自适应主动学习策略,通过聚类选择高关联性的特征,有效降低了特征维度,加快了评估进程。上述文献中的特征选择算法均与后续的电力系统TSA模型算法联系紧密,仅能用于对应的分类器中,扩展性较差。

基于上述分析,本文提出一种基于Fisher Score特征选择的电力系统TSA方法。Fisher Score算法通过计算各特征之间的类内距离和类间距离,能够有效筛选重要的电气特征,剔除冗余和噪声干扰信息,提升电力系统TSA的效率和准确性。该算法的主要流程为:首先,从电力系统运行的潮流量中人工挑选电气特征,并构建原始电气特征集^[11-13];然后,通过Fisher Score值评价原始电气特征集中所有电气特征的重要性;最后,按照电气特征的重要性排序,将得分较高的特征依次打包送入后续电力系统TSA模型,并通过实验检验Fisher Score特征选择算法对提升不同电力系统TSA模型性能的有效性。本文的主要创新点在于,通过Fisher Score算法实现对电力系统TSA中样本数据关键特征的提取以及对噪声干扰特征的剔除,该算法能够与多种类型的分类器结合,具有良好的扩展性。

1 基于Fisher Score特征选择的电力系统TSA方法

1.1 Fisher Score算法

Fisher Score是一种对数据样本特征进行评判的有效标准,传统的Fisher Score算法源自Fisher线性判别法,其主要思想是鉴别性能较强的特征,表现为在特征集空间中寻找类内距离尽可能小、类间距

离尽可能大的特征子集。文献[14]针对二分类问题,提出一种Fisher Score算法并将其作为特征选择准则,取得了很好的特征选择效果;文献[15]提出一种适用于多类别问题的Fisher Score算法。本文所描述的电力系统TSA问题属于一个二分类(稳定和不稳定)问题,因此,利用Fisher Score算法对电力系统TSA问题中的输入电气数据特征进行选择。

给定电气特征样本 $\mathbf{x}_i \in \mathbf{R}^m (i=1, 2, \dots, n)$, 其中 n 为样本总数, m 为样本的特征维数。假设所有电气特征样本中的稳定和不稳定样本数分别为 n_s 和 n_u , 利用Fisher Score算法计算得到样本的第 k 维特征得分 F_k 为:

$$F_k = \frac{(u_k^s - u_k)^2 + (u_k^u - u_k)^2}{\frac{1}{n_s - 1} \sum_{i=1}^{n_s} (x_{i,k}^s - u_k^s)^2 + \frac{1}{n_u - 1} \sum_{i=1}^{n_u} (x_{i,k}^u - u_k^u)^2} \quad (1)$$

式中: u_k 、 u_k^s 、 u_k^u 分别为样本的第 k 维特征在整个数据集中的均值、在稳定样本数据集中的均值、在不稳定样本数据集中的均值; $x_{i,k}^s$ 为第 i 个稳定样本数据的第 k 维特征的特征值; $x_{i,k}^u$ 为第 i 个不稳定样本数据的第 k 维特征的特征值。式(1)中等号右边的分子表示不同类样本数据的类间散度之和, 分母表示同一类样本数据的类内散度之和, 因此, 通过式(1)得到的Fisher Score值越大, 该维特征的判别力越强, 与电力系统暂态稳定分类问题的相关性也越高。通过式(1)能够很好地计算出原始电气特征集中每维电气特征的判别力, 但在电力系统暂态稳定二分类问题中无法考虑特征在稳定和不稳定样本间的一致性问题, 为此, 本文借鉴文献[16]的交叉系数思想, 如式(2)所示。

$$N_i = n_{si} + n_{ui} - n_{sui} \quad (2)$$

式中: N_i 为稳定样本和不稳定样本中包含特征样本 $\mathbf{x}_i$ 的样本数; n_{si} 为稳定样本中包含特征样本 $\mathbf{x}_i$ 的样本数; n_{ui} 为不稳定样本中包含特征样本 $\mathbf{x}_i$ 的样本数; n_{sui} 为稳定样本和不稳定样本中特征样本 $\mathbf{x}_i$ 取值相同的样本数。结合式(1)和式(2)能够更好地计算原始电气特征样本数据集中每维电气特征的判别力, 从而对不同电气特征与电力系统TSA分类问题之间的相关性大小进行判断。

1.2 基于Fisher Score特征选择的电力系统TSA流程

本文提出的基于Fisher Score特征选择的电力系统TSA方法的实施流程如附录A图A1所示, 具体步骤如下。

1) 通过电力系统仿真软件对所搭建的电力系统模型进行时域仿真, 通过设置不同的故障类型获取电力系统的暂态稳定数据。

2) 根据仿真得到的电力系统暂态稳定状态对所获取的数据进行归一化处理和状态标记, 区分稳定

和不稳定2种类别。将数据按照一定的比例分为训练数据集和测试数据集, 将训练数据集作为评估预测学习器的输入, 在训练得到电力系统暂态稳定的评估模型后, 再将测试数据集作为评估模型的输入, 得到评估模型的评估准确率。

3) 利用Fisher Score算法对原始输入数据集中的电气特征进行重要性计算, 将特征按照重要性从高到低排序, 并将特征按照顺序依次加入训练数据集中训练电力系统TSA模型, 得到不同输入特征维度下测试样本的评估准确率。

4) 通过比较所有特征维度下的准确率得到准确率的最大值以及对应的特征维度, 将准确率的最大值与未进行特征选择的准确率进行对比, 得到特征选择后训练模型的训练效果。

2 电力系统TSA模型

2.1 电力系统暂态稳定的问题描述

对于电力系统TSA问题, 样本数据为学习模型的输入, 将样本数据 $\mathbf{X}$ 表示为:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \quad (3)$$

在采用支持向量机(support vector machine, SVM)、深度神经网络等机器学习算法进行电力系统暂态稳定性评估时, 每个包含 m 维特征的样本数据均可看作是样本空间 $\mathbf{R}^m$ 中的一个向量。假设与样本数据对应的样本标签为 $\mathbf{Y}$, 如式(4)所示。由于在电力系统TSA问题中, 标签只有稳定和不稳定, 因此类别数为2。

$$\mathbf{Y} = [y_1, y_2, \dots, y_n]^T \quad (4)$$

式中: $y_i (i=1, 2, \dots, n)$ 为对应样本数据 $\mathbf{x}_i$ 的样本标签, 当其值为1时, 标记 $\mathbf{x}_i$ 稳定, 为0时标记 $\mathbf{x}_i$ 不稳定。

电力系统TSA问题就是寻找映射 $\mathbf{X} \rightarrow \mathbf{Y}$ 的问题, 通过分类器的训练模型得到样本数据和样本标签之间的映射关系, 如果能够找到合适的对应关系, 就能够达到较好的评估效果。

样本数据根据暂态稳定类别进行标记, 而暂态稳定类别根据转子功角的暂态稳定指数(transient stability index, TSI)进行判定, 该指数的表达式为:

$$C_{TSI} = \frac{360 - |\Delta\delta_{\max}|}{360 + |\Delta\delta_{\max}|} \quad (5)$$

式中: C_{TSI} 为TSI; $\Delta\delta_{\max}$ 为仿真时长内任意2台发电机的最大相对功角差。当 $C_{TSI} > 0$ 时, 判定样本数据稳定, 反之, 则判定样本数据不稳定。

2.2 输入特征的构建

利用机器学习方法进行电力系统TSA的关键问题之一是如何选择输入特征, 所选择的输入特征须物理意义明确且能够代表系统动态。文献[17]阐述

各种特征对电力系统暂态稳定性的重要程度,给出 TSA 问题中输入特征的选取方法。由于基于机器学习的电力系统 TSA 方法是一种数据驱动的学习方法,需要借助大量历史特征数据的学习和训练来构建评估模型,特征数据越全面,在构建模型时对电力系统的描述就越全面,因此,除发电机功角特征外,本文还选取了其他电气特征,以适应不同规模的电力系统。此外,为防止在仅使用发电机功角作为特征时因噪声干扰或采样设备故障而影响 TSA,基于现有文献^[18-19],本文从时间、空间及系统规模角度,通过大量仿真试验选取 26 维电力系统 TSA 原始特征构造特征集,各特征如表 1 所示,当不能准确采集发电机功角信号时,可以根据其他特征对暂态稳定进行准确评估。

表 1 TSA 原始特征

Table 1 Original features of TSA

特征序号	特征描述
1	发电机功角最大值与最小值之差
2	发电机功角平均值
3	发电机功角方差
4	发电机功角最大值
5	发电机功角最小值
6	发电机功角轨迹簇曲线的中心线
7	发电机速度偏差最大值与最小值之差
8	发电机速度偏差平均值
9	发电机速度偏差方差
10	发电机速度偏差最大值
11	发电机速度偏差最小值
12	发电机速度偏差轨迹簇曲线的中心线
13	发电机电磁功率平均值
14	发电机电磁功率方差
15	发电机无功功率平均值
16	发电机无功功率方差
17	发电机加速功率平均值
18	发电机加速功率方差
19	母线正序电压平均值
20	母线正序电压方差
21	母线有功负荷平均值
22	母线有功负荷方差
23	母线频率偏差平均值
24	母线频率偏差方差
25	发电机转子动能最大值
26	发电机转子动能最小值

表 1 中的原始特征主要为动态特征和系统特征,能准确反映故障对系统的冲击,且特征维度不随系统规模的变化而变化,能够满足大系统 TSA 的需求。为保证样本数据在时间上的全面性,在故障前、故障期间、故障后对所有特征均进行采样,对 3 个时间段的采样数据进行处理后得到最终的样本数据,在满足样本数据时间上的全面性要求的同时,避免了因故障持续时间不同而导致的暂态稳定状态的误判。此外,随着广域测量技术的发展,通过分布在各发电厂和变电站中的同步相量测量单元(phasor

measurement unit, PMU) 获取上述数据将越来越方便。

3 算例分析

3.1 新英格兰 39 节点系统

新英格兰 39 节点系统共包含 10 台发电机、39 条母线和 46 条线路,其中 39 号母线上的发电机是一台等值机,在本算例中,将其作为参考发电机。系统基准功率为 100 MV·A,额定频率为 60 Hz。

3.1.1 样本数据的构造

对于新英格兰 39 节点系统,样本的具体构造方法如附录 B 表 B1 所示。本文采用电力系统分析软件中的潮流计算及暂态稳定分析程序生成 4 800 个有效样本,其中稳定样本为 3 750 个,不稳定样本为 1 050 个。

对生成的 4 800 个样本进行标记,随机选取其中的 3 000 个样本作为训练样本,将剩余的 1 800 个样本作为测试样本。每个样本均以表 1 中的 26 维电气特征进行表示。

3.1.2 通过 Fisher Score 对特征进行重要性排序

利用第 1 节中的 Fisher Score 算法得到每维特征的得分,并对得分进行归一化处理,如图 1 所示。得分越接近 1 说明该维特征的重要性越高,反之重要性越低。

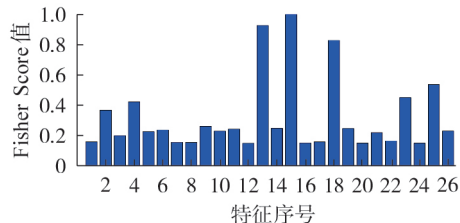


图 1 各特征的 Fisher Score 值

Fig.1 Fisher Score values of each feature

对于新英格兰 39 节点系统而言,由于系统规模较小,各发电机之间的联系相对密切,当在运行过程中系统的暂态稳定状态发生变化时,各发电机间的功角与功率变化较明显,因此,每台发电机的功角和功率与该系统的暂态稳定性具有较高的关联性。由图 1 可知,与发电机运行功角和功率相关的特征重要性较高,其他特征的重要性相对较低,例如发电机电磁功率平均值(第 13 维特征)、发电机无功功率平均值(第 15 维特征)、发电机加速功率方差(第 18 维特征)等特征的重要性较高,而母线正序电压方差(第 20 维特征)、母线频率偏差方差(第 24 维特征)等特征的重要性较低。这与文献^[12]中影响系统暂态稳定性的关键特征种类基本一致。

通过上述分析可以看出, Fisher Score 算法可以准确地筛选出与电力系统暂态稳定性相关性较高的特征。将 26 维特征按照重要性由高到低的顺序依

次加入样本数据的特征集中,在利用电力系统TSA分类模型进行学习训练之后,便可以获得在不同电气特征维度下电力系统暂态稳定的分类预测准确率。为了更全面地验证不同维度特征对电力系统TSA的影响程度,本文将SVM^[20-22]和深度置信网络^[23](deep belief network, DBN)2种分类器作为电力系统TSA模型,对经过Fisher Score特征选择后的数据进行学习训练和准确率预测测试,结果如图2所示。由图可知:按照得分由高到低的顺序依次将不同特征加入特征集后,随着加入重要特征数的增多,所有预测模型的准确率都呈现逐步提高的趋势;当加入的特征达到一定数量后,随着特征的继续加入,预测模型的准确率不再变化或者出现小幅下降。因此,经过Fisher Score特征选择后,仅利用重要性较高的少量特征便可以得到比较稳定的评估和预测结果,而重要性较低的特征对电力系统TSA模型的影响较小,甚至部分冗余特征的加入还会产生负面效应。

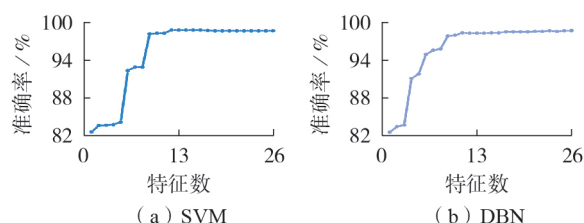


图2 采用不同分类器测试得到的准确率-特征维度折线图

Fig.2 Broken line chart of accuracy rate vs. feature dimension obtained by test of different classifiers

2种不同分类器模型在Fisher Score特征选择前后的准确率对比如表2所示。

表2 Fisher Score特征选择前后的准确率对比

Table 2 Comparison of accuracy rate between before and after Fisher Score feature selection

分类器	条件	特征维度	准确率 / %
SVM	特征选择前	26	98.67
	特征选择后	12	98.78
DBN	特征选择前	26	98.61
	特征选择后	23	98.72

由表2可知:在经过Fisher Score特征选择后,2种分类器模型的准确率均比特征选择前有所提高;不同分类器所需的关键特征数不同,SVM分类器利用前12维Fisher Score值最高的特征就可以得到最优结果,而DBN分类器需要利用前23维重要的特征才能取得最高的评估准确率;Fisher Score特征选择可以在一定程度上压缩原始特征的规模,减少冗余信息,提高分类器的训练和测试效率。

3.1.3 噪声干扰对Fisher Score效果的影响

在原有的26维特征中选取其中的第2、9、11、15、18维特征,对这些特征加入噪声干扰之后进行

噪声鲁棒性测试。在实际中,PMU可能受到的干扰包括随机噪声信号干扰、谐波信号干扰、直流信号干扰等,本文算例噪声信号选取最常见的随机噪声信号。由于在传统的TSA中一般将故障切除后的参数数据作为采样数据,且本文算例中故障切除后的采样数据较多,因此,设置噪声的具体干扰方式为:在仿真进行到第100个周期时,向电力系统中输入噪声信号,以发电机功角平均值为例,噪声干扰对时域曲线的影响如图3所示。由图可知,由于噪声的干扰,样本数据出现不规则的上下离散波动,这导致时域曲线由原本的平滑状态出现畸变,进而对该特征数据在数据空间中的分布产生负面影响,而Fisher Score算法是通过计算不同特征数据的类内距离与类间距离得到该特征的重要性得分的,因此,噪声干扰最终会影响特征选择算法对该特征的计算结果。

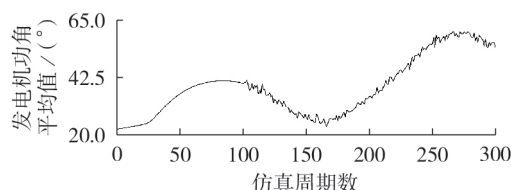


图3 噪声干扰对时域曲线的影响

Fig.3 Influence of noise interference on time domain curve

利用Fisher Score算法对噪声污染后的26维特征进行重要性得分计算,对得分进行归一化处理,得到加入噪声干扰前后不同特征的重要性对比结果,如图4所示。

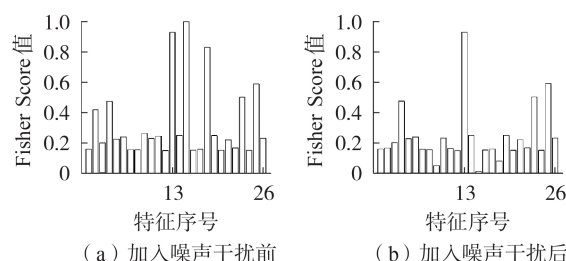


图4 加入噪声干扰前后的Fisher Score值对比

Fig.4 Comparison of Fisher Score values between before and after adding noise interference

由图4可知,相较于加入噪声干扰前,加入噪声干扰后的第2、9、11、15、18维特征的Fisher Score值明显降低。类间散度表示稳定与不稳定2类不同样本数据在特征空间分布中的离散程度,类间散度越大越有利于分类器模型对2类样本进行正确分类。类内散度表示同一类别样本数据在特征空间分布中的离散程度,类内散度越小,则同一类样本数据越聚集,这些样本也越容易被划分为同一类。由图3中的发电机功角平均值特征可知,加入噪声干扰后的

时域数据出现了较大的随机波动,这导致相同类别特征样本数据的类内散度增大,类间散度减小,进而使该维特征的Fisher Score值急剧下降,从而减小了该维特征对不同类别样本数据的分类判别作用,使其不被选择进入有效的特征子集。加入噪声干扰前后的Fisher Score值量化对比如附录B表B2所示。

对26维特征计算Fisher Score值,将特征按照重要性由高到低的顺序依次加入数据特征集中,并且利用分类器模型训练和测试电力系统TSA准确率,结果如附录B图B1所示。由图可知,经过Fisher Score特征选择后,随着重要性较高的特征的逐步加入,分类器模型的准确率均逐渐提高并达到最优结果,而在后期随着噪声干扰后重要性较低的特征的加入,部分分类器模型的准确率出现了一定程度的下降,这说明利用Fisher Score算法筛选含有噪声干扰的电气特征是十分必要的。

加入噪声干扰后Fisher Score特征选择前后的准确率对比如表3所示。由表3与表2可知:在原始特征中的部分特征受到噪声干扰后,无论采用哪种分类器,其评估准确率比受到噪声干扰前的均有所降低;受到噪声干扰的特征的重要性会降低,而通过利用Fisher Score算法筛选重要性较高的特征进行分类器训练后,评估准确率仍能维持在较高的水平。

表3 加入噪声干扰后Fisher Score特征选择前后的准确率对比

Table 3 Comparison of accuracy rate between before and after Fisher Score feature selection after adding noise interference

分类器	条件	特征维度	准确率 / %
SVM	特征选择前	26	98.56
	特征选择后	12	98.78
DBN	特征选择前	26	98.56
	特征选择后	23	98.72

通过分析发现各分类器的性能存在一定的差异:对于SVM分类器,由于该分类器通过由位于分类边界上的少数样本数据构成的支持向量决定最终的决策函数和分类结果,因此,可以在一定程度上避免输入特征的“维数灾”问题,但该分类器对于大规模数据以及多分类问题的分类性能稍有欠缺,且对参数设置和核函数的选择较为敏感,这使评估结果容易受到影响,经过Fisher Score特征选择后,该分类器模型的评估准确率提升幅度较小,但输入空间的维度得到了有效降低;对于DBN分类器,该分类器是由多层受限玻尔兹曼机(restricted Boltzmann machine, RBM)构成的,可通过多层RBM实现对数据的深度学习,从而对数据进行有效分类,由于DBN在学习过程中具有一定的特征提取能力,因而经过Fisher Score特征选择后仍需较多的特征来完成对

数据的学习任务,此时,输入空间的维度小幅降低,评估性能有所提升。

3.1.4 不同特征选择方法的性能对比

除Fisher Score算法外,本文还选取Relief、T检验、互信息法、主成分分析(principal component analysis, PCA)法这些特征选择方法进行对比分析。其中:Relief、T检验、互信息法均是利用各特征在分类方面的差异来完成对关键特征的筛选和噪声特征的剔除;PCA法是通过将多维特征映射到所需维度完成对高维数据的降维。为便于对比,本算例统一采用SVM作为上述特征选择方法的电力系统TSA模型。不同特征选择方法的准确率对比如表4所示。

表4 不同特征选择方法的准确率对比

Table 4 Comparison of accuracy rate among different feature selection methods

特征选择方法	准确率 / %
Relief	98.14
T检验	96.67
互信息法	96.56
PCA法	98.50
Fisher Score算法	98.78

由表4可知:计算Fisher Score得分之后,特征选择的有效性会大幅提升,有利于提高电力系统TSA的准确率;T检验、互信息法和Fisher Score算法均属于过滤式特征选择方法,但T检验和互信息法的准确率均低于本文的Fisher Score算法;PCA法通过对原始特征进行映射来降维,采用的不是特征的原始含义,并且其准确率也低于本文的Fisher Score算法。通过上述对比分析可知,Fisher Score算法在筛选电力系统暂态稳定关键电气量的特征以及提高电力系统TSA模型准确率方面具有较明显的优势。

3.2 IEEE 145节点系统

为进一步验证本文方法的有效性,采用更大规模的IEEE 145节点系统进行仿真验证。具体算例分析过程如附录C所示。可知:系统规模的不同导致暂态稳定的样本数据空间分布不同,这使Fisher Score算法对新英格兰39节点系统和IEEE 145节点系统中各类特征数据的类内距离与类间距离的计算结果出现差异,从而使得利用Fisher Score算法得到的IEEE 145节点系统的关键特征与新英格兰39节点系统的有所不同;但在剔除噪声干扰特征以及提升评估模型的评估性能方面,Fisher Score算法在IEEE 145节点系统中也能够取得较好的效果,这进一步验证了Fisher Score算法的有效性及良好的适应性。

4 结论

本文提出一种基于Fisher Score算法的电力系

统TSA方法,并在新英格兰39节点系统和IEEE 145节点系统中进行了仿真验证,得到如下结论:

1) Fisher Score算法在电气数据的特征选择方面具有良好的效果,对特征与电力系统TSA间的相关性大小有较高的判断力;

2) 对于随机噪声信号特征的干扰,利用Fisher Score算法能够很好地辨别与筛除,达到提高电力系统TSA模型准确率的效果;

3) 通过与其他特征选择方法的对比发现,Fisher Score算法能够更好地筛选电力系统TSA问题中的关键特征,以及更有效地提升预测模型的评估准确率。

本文通过利用Fisher Score特征选择对仿真所得到的样本数据特征集进行处理,有效减小了数据特征集的规模,并减少了数据特征中包含的冗余信息和噪声干扰,从而使得后续电力系统TSA模型的计算效率和准确率均得到大幅提升。此外,由于本文的Fisher Score算法与后续电力系统TSA分类模型之间相互独立,该算法可以方便地扩展用于各种不同的电力系统TSA模型。但是,目前在最优特征选择维度的设定方面仍需依赖人工经验,如何针对不同的电力系统TSA模型自适应地学习得到最优的特征维度,是笔者未来需要重点研究的内容。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 马宁宁,谢小荣,贺静波,等. 高比例新能源和电力电子设备电力系统的宽频振荡研究综述[J]. 中国电机工程学报,2020,40(15):4720-4732.
MA Ningning, XIE Xiaorong, HE Jingbo, et al. Review of wide-band oscillation in renewable and power electronics highly integrated power systems[J]. Proceedings of the CSEE, 2020, 40(15):4720-4732.
- [2] ZADKHAH S, JATSKOVICH J, VAAHEDI E. A multi-decomposition approach for accelerated time-domain simulation of transient stability problems[J]. IEEE Transactions on Power Systems, 2015, 30(5):2301-2311.
- [3] ZHU Q M, CHEN J F, ZHU L, et al. A deep end-to-end model for transient stability assessment with PMU data[J]. IEEE Access, 2018, 6:65474-65487.
- [4] 周艳真,吴俊勇,于之虹,等. 基于转子角轨迹簇特征的电力系统暂态稳定评估[J]. 电网技术,2016,40(5):1482-1487.
ZHOU Yanzhen, WU Junyong, YU Zhihong, et al. Power system transient stability assessment based on cluster features of rotor angle trajectories[J]. Power System Technology, 2016, 40(5):1482-1487.
- [5] 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 features[J]. Proceedings of the CSEE, 2021, 41(22):7596-7608.
- [6] 赵高尚,刘道伟,陈树勇,等. 基于尺度不变特征转换的暂态稳定边界特征提取[J]. 中国电机工程学报,2019,39(增刊1):84-94.
ZHAO Gaoshang, LIU Daowei, CHEN Shuyong, et al. Research on feature extraction of transient stability boundary based on scale invariant feature transform algorithm[J]. Proceedings of the CSEE, 2019, 39(Supplement 1):84-94.
- [7] LÓPEZ M, VALERO S, SENABRE C, et al. Analysis of the influence of meteorological variables on real-time short-term load forecasting in Balearic Islands[C]//2017 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering. Cadiz, Spain: IEEE, 2017:10-15.
- [8] 郭春明,任继红. 基于XGBoost-EE的电力系统暂态稳定评估方法[J]. 电力自动化设备,2021,41(2):138-143,152.
WU Chunming, REN Jihong. Power system transient stability assessment method based on XGBoost-EE[J]. Electric Power Automation Equipment, 2021, 41(2):138-143, 152.
- [9] 王怀远,陈启凡. 基于堆叠变分自动编码器的电力系统暂态稳定评估方法[J]. 电力自动化设备,2019,39(12):1-6.
WANG Huaiyuan, CHEN Qifan. Evaluation method of power system transient stability based on stacked variational automatic encoder[J]. Electric Power Automation Equipment, 2019, 39(12):1-6.
- [10] 卢东昊,王莉,张少凡,等. 基于聚类自适应主动学习的电力系统暂态稳定评估[J]. 电力自动化设备,2021,41(7):176-181.
LU Donghao, WANG Li, ZHANG Shaofan, et al. Transient stability assessment of power system based on clustering adaptive active learning[J]. Electric Power Automation Equipment, 2021, 41(7):176-181.
- [11] 谢彦祥,刘天琪,苏学能. Hadoop架构下基于分布式粒子群算法的暂态稳定评估特征量选择[J]. 电网技术,2018,42(12):4107-4115.
XIE Yanxiang, LIU Tianqi, SU Xueneng. Feature selection for transient stability assessment applying distributed PSO algorithm based on Hadoop architecture[J]. Power System Technology, 2018, 42(12):4107-4115.
- [12] 黄丹,孙华东,周勤勇,等. 基于响应轨迹最大李雅普诺夫指数动态特征的暂态稳定在线监测[J]. 电力自动化设备,2020,40(4):48-55.
HUANG Dan, SUN Huadong, ZHOU Qinyong, et al. Online-monitoring of transient stability based on largest Lyapunov exponent dynamic characteristics of response trajectory[J]. Electric Power Automation Equipment, 2020, 40(4):48-55.
- [13] 郑超,孙华东,陈怡君. 基于广域支路响应特征的失稳预判与紧急控制[J]. 中国电机工程学报,2021,41(17):5866-5877.
ZHENG Chao, SUN Huadong, CHEN Yijun. Instability pre-judge and emergency control based on the response characteristic of wide-area branches[J]. Proceedings of the CSEE, 2021, 41(17):5866-5877.
- [14] CHEN Y W, LIN C J. Combining SVMs with various feature selection strategies[M]//GUYON I, NIKRAVESH M, GUNN S, et al. Feature extraction: foundations and applications. Berlin, Germany: Springer, 2006:315-324.
- [15] 谢娟英,王春霞,蒋帅,等. 基于改进的F-score与支持向量机的特征选择方法[J]. 计算机应用,2010,30(4):993-996.
XIE Juanying, WANG Chunxia, JIANG Shuai, et al. Feature selection method combining improved F-score and support vector machine[J]. Journal of Computer Applications, 2010, 30(4):993-996.
- [16] TAO P, YI H, WEI C, et al. A method based on weighted F-score and SVM for feature selection[C]//2013 25th Chinese Control and Decision Conference (CCDC). Guiyang, China:

- IEEE,2013:4287-4290.
- [17] 朱林,张健,陈达,等. 面向暂态电压稳定评估的卷积神经网络输入特征构建方法[J]. 电力系统自动化,2022,46(1):85-93.
ZHU Lin,ZHANG Jian,CHEN Da,et al. Construction method for input features of convolutional neural network for transient voltage stability assessment[J]. Automation of Electric Power Systems,2022,46(1):85-93.
- [18] 黄辉,舒乃秋,李自品,等. 基于信息融合技术的电力系统暂态稳定评估[J]. 中国电机工程学报,2007,27(16):19-23.
HUANG Hui,SHU Naiqiu,LI Zipin,et al. Power system transient stability assessment based on information fusion technology[J]. Proceedings of the CSEE,2007,27(16):19-23.
- [19] 顾雪平,曹绍杰. 神经网络在暂态稳定评估中应用的研究述评[J]. 华北电力大学学报,2003,30(4):11-16.
GU Xueping,CAO Shaojie. Research overview of neural network applications to transient stability assessment of power systems[J]. Journal of North China Electric Power University,2003,30(4):11-16.
- [20] ZHANG W L,HU W,MIN Y,et al. A novel stability classifier based on reformed support vector machines for online stability assessment[C]//2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference(APPEEC). Brisbane,QLD, Australia:IEEE,2015:1-5.
- [21] 鲍颜红,冯长有,任先成,等. 基于支持向量机的在线暂态稳定故障筛选[J]. 电力系统自动化,2019,43(22):52-58.
- BAO Yanhong,FENG Changyou,REN Xiancheng,et al. On-line transient stability fault screening based on support vector machine[J]. Automation of Electric Power Systems,2019,43(22):52-58.
- [22] 黄彦浩,于之虹,史东宇,等. 基于海量在线历史数据的大电网快速判稳策略[J]. 中国电机工程学报,2016,36(3):596-603.
HUANG Yanhao,YU Zhihong,SHI Dongyu,et al. Strategy of huge electric power system stability quick judgment based on massive historical online data[J]. Proceedings of the CSEE,2016,36(3):596-603.
- [23] 史鹤欢,许悦雷,杨志军,等. 基于深度置信网络的目标识别方法[J]. 计算机应用,2014,34(11):3314-3317,3322.
SHI Hehuan,XU Yuelei,YANG Zhijun,et al. Target recognition method based on deep belief network[J]. Journal of Computer Applications,2014,34(11):3314-3317,3322.

作者简介:



李 鹏(1987—),男,副教授,博士,主要研究方向为机器学习和人工智能理论方法及其在电力系统分析和电气设备故障检测中的应用等(E-mail:lipeng@upc.edu.cn)。

(编辑 王锦秀)

Transient stability assessment method for power system based on Fisher Score feature selection

LI Peng,DONG Xinjian,MENG Qingwei,CHEN Jiming

(Department of Electrical Engineering,College of New Energy,China University of Petroleum,Qingdao 266580,China)

Abstract: In order to solve the problems of different correlation degrees between different electrical input features and transient stability of power system and the obvious decrease of assessment accuracy rate when the input features are interfered,a transient stability assessment method for power system based on Fisher Score feature selection is proposed. A Fisher Score value calculation scheme of sample features for transient stability assessment binary classification problem of power system is designed. Fisher Score value ranking is used to effectively distinguish important features from redundant features,noise features from non-noise features. The selected electrical features are input into different machine learning models for training and assessment. The simulative results of New England 39-bus system and IEEE 145-bus system show that the proposed feature selection scheme can effectively screen out the features of high importance in the transient stability assessment of power system,and improve the prediction performance of the assessment model.

Key words: electric power systems;transient stability assessment;feature selection;Fisher Score algorithm

附录 A

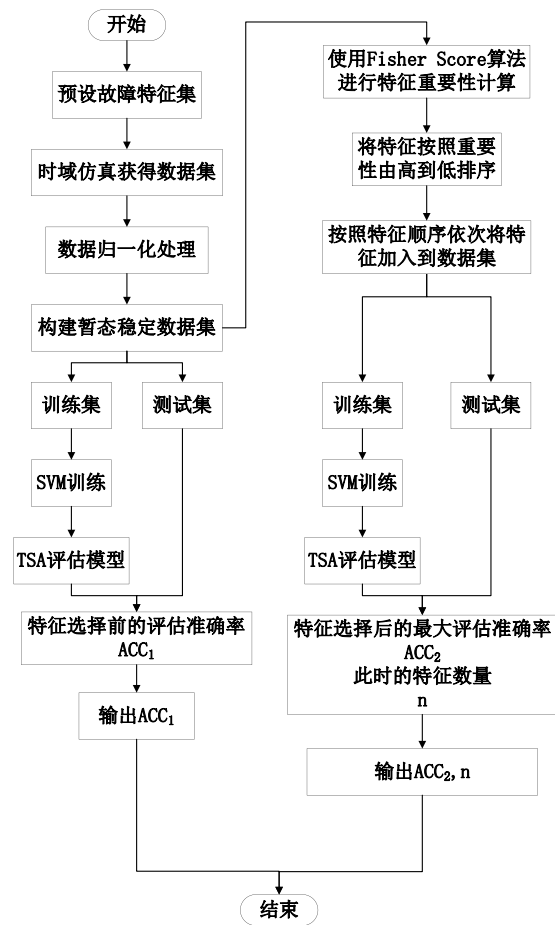


图 A1 基于 Fisher Score 特征选择的电力系统 TSA 方法流程图

Fig.A1 Flowchart of power system TSA method based on Fisher Score feature selection

附录 B

表 B1 新英格兰 39 节点系统仿真算例的样本数据构造方法
Table B1 Sample data construction method of New England 39-bus system simulation example

参数		参数值
发电机模型		二阶经典模型
负荷模型		恒阻抗模型
负荷水平		85%、90%、……、120%（共计 8 个）
故障位置		距离线路始端 0、20%、50%、80%（共计 4 个）
故障持续时间		1、3、6、9、12 个周期（共计 5 个）
故障类型		三相永久性短路
采样设置	采样频率	30 Hz
	故障前采样	由第 20 周波起采样持续 10 个周波
	故障期间采样	由第 30 周波起视故障持续时间设置采样时长
	故障后采样	由第 50 周波起采样持续 100 个周波； 由第 180 周波起采样持续 120 个周波

表 B2 第 2、9、11、15、18 维特征加入噪声干扰前后的 FisherScore 值
Table B2 Fisher Score values before and after No. 2, 9, 11, 15 and 18 features adding noise interference

特征维度	FisherScore 值	
	加入噪声干扰前	加入噪声干扰后
2	0.418	0.165
9	0.264	0.050
11	0.245	0.162
15	1.000	0.010
18	0.830	0.080

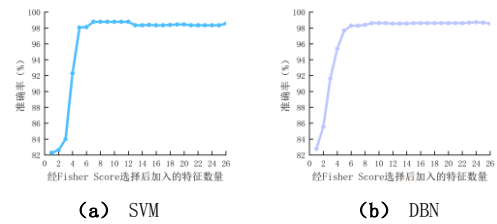


图 B1 对第 2、9、11、15、18 维特征进行噪声污染后采用不同分类器测试得到的准确率-特征维度折线图
Fig.B1 Broken line chart of Accuracy rate vs. feature dimension obtained by tested of different classifiers after No. 2, 9, 11, 15 and 18 features polluted by noise

附录 C

算例 2 IEEE 145 节点系统仿真实验验证

C1 算例系统介绍

为了进一步验证本文方法的有效性,本节采用了更大规模的 IEEE 145 节点系统进行实验验证。IEEE145 节点系统由 50 台发电机、145 条母线和 453 条传输线构成^[39]。50 台发电机中,1—6 号发电机和 23 号发电机为 6 阶模型,其他 43 台发电机为经典模型,负荷模型为恒阻抗模型。

C2 样本数据构造

对于 IEEE145 节点系统,共考虑 100%、105%、...、115% 共计 4 个负荷水平,相应改变发电机的出力保证系统功率平衡且各母线电压维持在 $0.95 \sim 1.05$ p.u. 的合理范围内,将故障设置在系统的输电线路,故障位置分别设置于距离线路时段 0、20%、50%、80% 处。故障类型设置为三相永久性短路,故障持续时间分别考虑 1、3、6、9、12 个周期,仿真时长为 300 个周期,暂态稳定判据同新英格兰 39 节点测试系统,仿真共生成 8 340 个样本,其中稳定样本 4 800 个,不稳定样本 3 540 个。随机选择 6 800 个样本进行训练,剩余 1 540 个样本用于测试。每一组数据样本仍按照表 1 所示构建原始 26 维电气特征量进行表示。

C3 通过 Fisher Score 对特征进行重要性排序

通过 Fisher Score 计算得到的 IEEE145 节点系统的各个特征量的重要性分数如图 C1 所示。

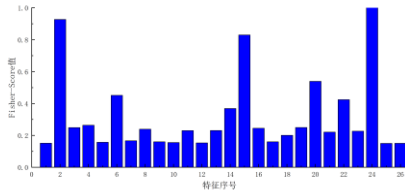


图 C1 各特征的 Fisher Score 值
Fig.C1 Fisher Score values of each feature

由于所选择系统的结构不同,不同系统中各个特征量对于暂态稳定性的重要程度也出现变化,所以同一特征在新英格兰 39 节点系统和 IEEE 145 节点系统对系统暂态稳定性的关联程度有所差异。通过分析图 7 中的各项特征量的重要性分数可以看出,发电机功角均值(第 2 维特征)、母线有功负荷平均值(第 21 维特征)等特征量的重要性排名较为靠前,而发电机转子动能最小值(第 25 维特征)、发电机转子动能最大值(第 26 维特征)等特征量的重要性排名较为靠后。针对 IEEE 145 节点系统而言,首先由于系统中发电机数量比较多,系统本身在运行过程中的抗

干扰能力较强,所以针对该系统的暂态稳定性,除各发电机之间的相对功角具有较高关联性外,系统中各种潮流量特征的变化对系统暂态稳定性的影响也更大;其次,由于 IEEE 145 节点系统规模更大,评估模型需要更多的动态特征来对系统加以描述。故而相比较于 IEEE 39 节点系统, Fisher Score 算法在对特征进行选择时对系统中的动态特征赋予了更高的分值,而对转子动能特征的分值相对减少。另一方面,传统的暂态稳定分析方法是通过分析发电机的转子动能等能量特征得出电力系统暂态稳定的状态,而基于机器学习的暂态稳定评估方法,主要是通过分析各项特征数据的数据空间分布,从而达到对暂态稳定的准确评估。而 Fisher Score 算法,是通过计算数据样本分布空间中的类内距离和类间距离分析得到每种特征量的得分大小以及对暂态稳定状态分类的作用大小。例如针对发电机转子动能最大值这一特征,由于不同系统的数据空间分布不同,在 IEEE 145 节点系统的数据空间中,转子动能的数据与其它特征数据相比 Fisher Score 值较低,意味着其在数据样本空间分布中的类内类间散度判别优势不如其他特征更为明显;而在 IEEE 39 节点系统的数据样本空间中转子动能特征的数据分布可以很好地对其系统暂态稳定性进行判别,所以在不同系统中,经过 Fisher 特征选择算法选择得出的特征子集略有不同。针对 IEEE 39 节点系统和 IEEE 145 节点系统母线有功负荷平均值的重要性得分出现变化,通过对仿真实验结果进行分析,是由于构建样本时两个系统所考虑的负荷水平不完全相同所导致的。需要说明的是, Fisher Score 值的高低是一个相对的概念,得分越高的特征是指在当前场景条件下作用更明显,但并不是 Fisher Score 值很低的特征在电力系统暂态稳定评估中就是完全没有任何作用的, Fisher Score 特征选择方法的主要优势是可以在各种不同的应用场景和系统中将尽可能少数且有效的数据特征筛选出来,从而快速准确地进行电力系统暂态稳定评估。

将处理好的特征量按照与新英格兰 39 节点测试系统相同步骤进行仿真实验,验证 Fisher Score 特征选择对于大规模电力系统的有效性。

通过使用 4.3 节中的 2 种分类器,进一步验证各个特征量对于 IEEE145 节点系统暂态稳定的影响。

使用 2 种分类器作为电力系统暂态稳定评估模型得到 Fisher Score 特征选择后的样本数据的评估结果如图 C2 所示。

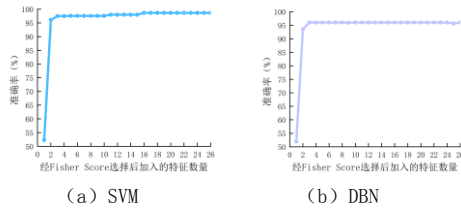


图 C2 采用不同分类器测试得到的准确率-特征维度折线图

Fig.C2 Broken line chart of accuracy vs. feature dimension obtained by test of different classifiers

通过图 B2 可以看出,准确率-特征数量的曲线变化趋势与新英格兰 39 节点系统的测试结果基本一致,可以得出相同结论。2 种分类器的量化对比结果如表 C1 所示。

表 C1 Fisher Score 特征选择前后准确率对比

Table C1 Comparison of accuracy between before and after Fisher Score feature selection

分类器	条件	特征维度	准确率/%
SVM	特征选择前	26	97.53
	特征选择后	13	98.89
DBN	特征选择前	26	95.65
	特征选择后	10	96.04

通过表 C1 中结果对比可知,由于 IEEE145 节点系统与新英格兰 39 节点系统的结构有所差异,所以相同特征量在形成评估模型后评估准确率会有所不同。但经过 Fisher Score 特征选择后,各个分类器形成的评估模型在准确率上均有不同程度的提升,并且选择出了相对应的关键特征量。

4 挑选部分特征进行噪声污染测试 Fisher Score 效果

为进一步测试Fisher Score特征选择的有效性,从原始特征集中挑选出 2、6、15、20、24 号特征进行噪声污染,使特征量具有一定的随机性,噪声信号的设置同 IEEE 39 节点系统一致,其对时域信号的干扰如图 C3 所示。

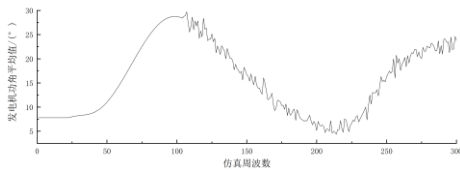
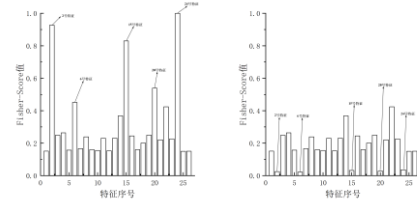


图 C3 噪声干扰时域图

Fig.C3Time domain diagram of noise interferes

然后通过 Fisher Score 特征选择计算各个特征量的重要性分数,结果如图 C4 所示。



(a) 噪声干扰前 (b) 噪声干扰后

图 C4 加入噪声干扰前后的 Fisher Score 值对比

Fig.C4 Comparison of Fisher Score values between before and after adding noise interference

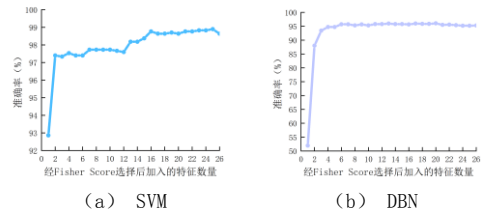
通过图 C4 中的对比结果可以看出,随机加入噪声信号后的第 2、6、15、20、24 维特征与未受到污染之前相比,其 Fisher Score 的重要性得分有了明显降低,,故而可以得出与 IEEE 39 节点系统添加噪声干扰时相同的结论。添加噪声后各特征具体的 Fisher Score 值的量化对比如表 C2 所示。

表 C2 加入噪声干扰前后的 FisherScore 值

Table C2 Fisher-score values before and after adding noise interference

特征序号	Fisher-Score 值	
	加入噪声干扰前	加入噪声干扰后
2	0.927	0.025
6	0.451	0.022
15	0.831	0.032
20	0.540	0.028
24	1.000	0.036

将经过随机信号干扰的 26 维特征在计算其 Fisher Score 后按照重要性分数由高到低依次加入到数据特征集中,并利用分类器模型训练和测试电力系统暂态稳定评估准确率,得到准确率-特征数量折线图如图 C5 所示。



(a) SVM (b) DBN

图 C5 加入噪声干扰后采用不同分类器测试得到的准确率-特征维度折线图

Fig.C5 Broken line chart of accuracy vs. feature dimension obtained by test of different classifiers after adding noise interference

通过图 B5 可以看出,在将特征量逐一加入到评估模型的过程中,各个分类器的评估准确率是逐渐上升的,等到达最高准确率的时候由于受到噪声污染的电器特征量的加入,使得评估模型的评估性能出现了不同程度的下降,最终评估结果和选择后的特征量如表 C3 所示。

表 C3 加入噪声干扰后 Fisher Score 特征选择前后的准确率对比
Table C3 Comparison of accuracy between before and after Feature
Scorefeatureselection after adding noise interference

分类器	条件	特征维度	准确率/%
SVM	特征选择前	26	96.64
	特征选择后	15	98.63
DBN	特征选择前	26	95.26
	特征选择后	20	96.04

根据表 C3 与表 C1 的对比可知，当原始数据特征中有部分特征受到噪声信号干扰时，无论采用哪种分类器，其评估准确率相对于数据特征受到噪声干扰前都有所较低。而通过使用 Fisher Score 特征选择方法，受到随机噪声干扰的特征的重要性排名会降低，因此，通过筛选出前部重要性排名较高的数据特征进行分类器训练后，其评估预测的准确率仍然能够维持在之前的较高水平。并且对比表 C3 与表 3 的实验结果，四种分类器性能在两个算例中表现基本一致，对于 IEEE 145 节点系统，各分类器经特征选择后也可有效降低输入空间的维数，提升评估准确率。