

数据驱动的含IIDG配电网短路电流计算多输出模型

叶睿恺,王慧芳,张 森,张亦翔

(浙江大学 电气工程学院,浙江 杭州 310027)

摘要:“双碳”目标下大量逆变型分布式电源(IIDG)接入配电网,使基于机理建模的配电网短路电流计算中,计算速度与准确性间的矛盾日益突出。数据驱动建模方法已被证实能有效解决两者间的矛盾,但采用单输出模型难以满足实际应用中要输出多个计算点的需求,容易产生模型数量问题。针对上述问题,提出了配电网短路电流多输出回归计算模型与计算方法。对问题转化与算法适应这2类多输出模型进行了分析对比,提出基于多目标回归模型融合方法、回归链方法等问题转化方法以及神经网络多输出方法,适合解决含IIDG配电网短路电流计算问题。分析了输入特征选择、性能评价指标、计算流程以及超参数寻优方法等关键问题。算例表明多输出模型能同时满足计算准确性和计算速度要求,性能强于单输出模型,且避免了模型数量问题。

关键词:短路电流计算;多输出回归;逆变型分布式电源;配电网;数据驱动建模

中图分类号:TM 713

文献标志码:A

DOI:10.16081/j.epae.202205015

0 引言

我国“双碳”战略目标促使逆变型分布式电源IIDG(Inverter Interfaced Distributed Generator)在配电网中逐渐呈现高占比的发展趋势^[1]。与传统同步电机电源不同,IIDG具有强非线性的故障特性,不仅要满足低压穿越要求^[2],还受故障前运行特性、接入位置、控制策略等因素影响。大量IIDG的接入使得传统短路电流计算方法难以直接应用,迫切需要对含IIDG配电网短路电流计算方法进行研究^[3]。

针对IIDG故障特性以及短路电流计算方法,国内外学者已开展了很多研究。文献[4]详细介绍了IIDG的低压穿越控制策略,并给出了短路电流的表达式;文献[5-6]根据不同目标的控制策略分析了不对称故障下IIDG的电流输出特性;文献[7]分析了IIDG并网后的故障特性,得到了对称故障与不对称故障下短路电流计算模型;文献[8]分析了IIDG并网后对配电网的影响,给出了含少量IIDG配电网短路电流计算通用模型;文献[9]将IIDG分为故障点上游和故障点下游并分别建立故障等效模型,采用基于叠加原理的迭代短路电流计算方法。上述研究均是基于机理的物理建模,不同控制策略下IIDG模型不同,不同故障类型下不同的短路电流计算方法都需要迭代计算。当IIDG大量接入或网络规模增大时,上述方法因迭代计算耗时必将大增,甚至可能出现不收敛的情况。为提升计算速度,有学者对

IIDG模型做极简处理,例如用1.2~2倍额定电流来表征,但这必将牺牲计算准确性^[10]。因此现有物理建模方法应用于IIDG占比高的大型配电网时,会出现计算快速性与准确性之间的矛盾。

为了解决该矛盾,文献[11]基于机器学习思想提出了机理与数据融合驱动的含IIDG配电网短路电流计算方法,并与详细的物理建模方法进行对比,结果表明文献[11]所提方法的计算误差接近物理建模方法的计算误差或更小,且计算速度快。同时,电网规模、IIDG接入数量对该方法的计算性能影响较小^[12]。然而,该方法使用单输出的学习模型,当需要计算配电网中多个测量点电流值时,必须针对每个测量点训练各自的学习模型,因此存在模型数量多的问题。

为解决模型数量问题,本文提出了含IIDG配电网短路电流计算多输出模型。对不同的多输出模型进行分析与对比,并根据获得的配电网运行状态的实际条件,选择2种包含运行方式和故障信息的输入特征,分别进行模型训练,最终实现了同时输出配电网所有支路短路电流的目标。

1 多输出回归模型选择

传统机器学习的目标多为建立单输出模型,即采用训练完成的模型对特定目标进行回归计算。实际问题中往往需要同时解决多个任务,且多个任务之间存在不同的相关性,因此需要研究多输出回归计算。根据多输出回归的子任务关联方式,大致可分为问题转化和算法适应方法2类^[13]。

1.1 问题转化方法

多输出任务目标是建立 M 个输入与 D 个输出之间的映射关系。问题转化的核心思想是将相应的映

收稿日期:2021-08-25;修回日期:2022-03-22

在线出版日期:2022-04-08

基金项目:国家自然科学基金联合基金资助项目(U2166204)

Project supported by the Joint Funds of the National Natural Science Foundation of China(U2166204)

射关系转化为 D 个单输出的形式,然后将 D 个单输出模型拼合成1个多输出模型,达到同时输出多个预测值的目的。问题转化的算法包括单目标方法ST(Single-Target method)、多目标回归模型融合MTRS(Multi-Target Regressor Stacking)方法和回归链RC(Regressor Chain)方法等。

设有 N 组样本, X^l 为第 l 组的输入, Y^l 为第 l 组对应的输出,即第 l 组样本的输入、输出为:

$$\begin{cases} X^l=(x_1^l, \dots, x_i^l, \dots, x_M^l) \\ Y^l=(y_1^l, \dots, y_j^l, \dots, y_D^l) \end{cases} \quad l \in \{1, 2, \dots, N\} \quad (1)$$

式中: x_i^l 为 X^l 中第 i 个输入特征, $i \in \{1, 2, \dots, M\}$; y_j^l 为 Y^l 中第 j 个输出值, $j \in \{1, 2, \dots, D\}$ 。

ST是最直接的将问题转化成多输出模型的方法,分别建立所有输出的映射,对于 D 个输出需要建立 D 个单输出模型,由每个单输出模型负责单个输出预测值。第 j 个输出的训练集 D_{ST-j} 为:

$$D_{ST-j}=\left\{\left(X^1, y_j^1\right), \dots, \left(X^l, y_j^l\right), \dots, \left(X^N, y_j^N\right)\right\} \quad (2)$$

ST中每个输出均为独立输出,参数之间不会相互影响,充分挖掘了全部输入向量对特定输出量的影响,单项预测准确率较高,但忽略了不同输出之间的潜在关系,因此适用于输出任务关联不密切的问题。

为了利用输出量之间的潜在关系,MTRS方法在ST的基础上做出改进,建立一个由2个阶段组成的多输出模型。MTRS将ST作为第一阶段,获得第一阶段输出量 $\hat{Y}^l=(\hat{y}_1^l, \dots, \hat{y}_j^l, \dots, \hat{y}_D^l)$,将 $\hat{Y}^l$ 和原始输入量 X^l 合并为新的输入 X^{*l} ,即:

$$X^{*l}=(x_1^l, \dots, x_i^l, \dots, x_M^l, \hat{y}_1^l, \dots, \hat{y}_j^l, \dots, \hat{y}_D^l) \quad (3)$$

MTRS的第二阶段方法使用新输入量 X^{*l} 和原始输出 Y^l 作为 D 个输出模型的训练集,即第 j 个输出的训练集 D_{MTRS-j} 为:

$$D_{MTRS-j}=\left\{\left(X^{*1}, y_j^1\right), \dots, \left(X^{*l}, y_j^l\right), \dots, \left(X^{*N}, y_j^N\right)\right\} \quad (4)$$

第二阶段通过将第一阶段的输出预测值拼合作为训练集的输入,在一定程度上考虑了多个输出之间的潜在关系,对问题转化预测值偏差进行了一定的修正,可以提升预测准确性。但是,MTRS方法本质上与ST类似,都是直接考虑输入和输出之间的映射,对于输出之间的潜在关系的考虑并不全面。此外,MTRS方法需要进行2次ST计算,使得模型偏大、计算速度慢。

RC方法是基于马尔可夫随机链的思想,根据确定的链顺序对每个输出建立独立的回归模型。当有 D 个输出值时,假设指定链顺序 C 为:

$$C=(y_1, y_2, \dots, y_D) \quad (5)$$

依据链顺序 C 排列获得模型顺序。第1个模型仅使用 X^l 输出第1个预测值 y_1^l ,将 y_1^l 加入 X^l 得到

$X_2^{*l}=(X^l, y_1^l)$ 用于输出预测值 y_2^l ,将 y_2^l 加入 X_2^{*l} 得到 $X_3^{*l}=(X^l, y_1^l, y_2^l)$ 用于输出预测值 y_3^l ,以此类推。因此 y_1^l 之后的模型输入使用原始输入和链顺序之前的标签值,充分考虑前序输出对当前输出值的影响,即采用链顺序 C 将第 j 个输出的输入特征划分为:

$$X_j^{*l}=(X^l, y_1^l, y_2^l, \dots, y_{j-1}^l) \quad 1 < j \leq D \quad (6)$$

因此,第 j 个输出的训练集 D_{RC-j} 为:

$$D_{RC-j}=\left\{\left(X_j^{*1}, y_j^1\right), \dots, \left(X_j^{*l}, y_j^l\right), \dots, \left(X_j^{*N}, y_j^N\right)\right\} \quad (7)$$

最终将 D 个预测值一起输出。由上述过程可见,RC方法将不同输出考虑为不同状态,将不同状态以输入的方式挖掘多个输出之间的相关性,因此考虑了不同输出之间的关系。针对不同的实际问题,应根据输出任务间的相关性选择不同的问题转化方法,达到最佳的多输出性能。含IIDG配电网短路电流计算问题中,由于IIDG与系统电源提供的短路电流相位往往不同^[8],各支路稳态电流的幅值之间不再显式地满足基尔霍夫电流定律,但电流相量之间还是满足基尔霍夫电流定律的,即输出之间存在一定的相关性,因此将其他模型的输出量纳入输入的MTRS和RC方法这2种多输出模型方法适用于配电网短路电流多输出计算,可充分挖掘不同支路电流间的潜在关系。

由于MTRS和RC方法是将单输出模型拼合为多输出模型,因此方法的性能与作为基学习器的单输出模型性能密切相关。配电网短路电流计算由于输入特征数量较多,使模型复杂程度增大;同时输出电流数量多且有部分数值相近,因而存在一定的过拟合风险。为此本文选择近年来在各大比赛中性能表现优秀的基学习器——极限梯度提升机XGBoost(eXtreme Gradient Boosting)^[14]和轻量梯度提升机LightGBM(Light Gradient Boosting Machine)^[15]。XGBoost和LightGBM都是梯度提升决策树GBDT(Gradient Boosted Decision Tree)的改进与高效实现,其中XGBoost采用的是按层生长策略,能够同时分裂同一层的叶子,从而进行多线程优化,可有效降低过拟合风险;LightGBM采用带有深度限制的按叶子生长策略,每次从当前所有叶子中找到分裂增益最大的叶子进行分裂,在分裂次数相同的情况下,该策略可以降低更多的误差。此外,LightGBM使用了基于梯度的单边采样对输入进行降维以降低模型复杂程度,因此其可以在保证计算准确度的前提下降低模型复杂程度,使得模型计算速度更快。

但是,MTRS方法每进行一组样本训练都需要更新一次输入特征,RC方法则需要更新 $D-1$ 次输入特征,这可能导致LightGBM速度慢于XGBoost。原因是LightGBM需要对新输入特征使用互斥特征捆绑EFB(Exclusive Feature Bundling),当输入特征不

变或变化次数少时,EFB能大幅提升算法性能,当输入特征变化次数多时,EFB有可能减慢速度。因此,在MTRS方法下LightGBM的速度可能比XGBoost快,在RC方法下LightGBM的速度可能比XGBoost慢。

综上所述,含IIDG配电网短路电流计算多输出模型采用问题转化方法时,分别采用XGBoost和LightGBM作为基学习器,并对比基于MTRS和RC方法的多输出模型性能。

1.2 算法适应方法

算法适应方法是将单输出模型根据实际问题进行修改,用1个模型同时输出多个预测值。算法适应方法的输出映射不仅需要考虑输入变量,更需要考虑不同输出量之间的关系,以此提升多个预测值的准确性。

算法适应方法包含统计学方法、多输出支持向量机、多输出回归树、核函数方法、神经网络等多种方法。算法在修改之前是适用于单输出模型的,针对不同的实际问题,需要修改算法中的输入、模型结构、输出标签、损失函数等参数,不同的算法适应方法修改方式并不完全相同。本文采用应用较为广泛的神经网络作为多输出回归预测的基础算法。

神经网络一般是指全连接的人工神经网络ANN(Artificial Neural Network),它将输入层的神经元全部连接至隐藏层,利用所有输入特征计算得到输出值,因此是利用输入特征最充分、计算结果较为准确的模型。然而,全连接人工神经网络对训练样本数量要求高、模型计算量大。卷积神经网络CNN(Convolutional Neural Network)可以解决以上问题,其卷积层可以通过更少的计算步骤实现类似效果,对于复杂模型更有优势。CNN的卷积层通过滑窗实现输入特征的局部感知,即相邻输入数据可以为输出提供更高价值。电网具有基尔霍夫电流、电压定律,因此将输入的电气量特征相邻排列,可使CNN的卷积核获得有益信息,进而提升短路电流计算准确性。

本文选用ANN和CNN作为算法适应方法的代表算法,验证算法适应多输出模型应用于含IIDG配电网短路电流计算的有效性。

2 多输出模型的短路电流计算方法

2.1 输入特征选择

目前,IIDG高占比的配电网示范工程已配置有微型同步相量测量单元 μ PMU(Micro-synchronous Phasor Measurement Unit),可直接获得电压幅值、电压相位、有功、无功等数据,这些数据量反映了配电网的运行方式。然而,大部分配电网不具备上述条件,但一般在各支路上安装有电流互感器,能获得支路电流的有效值,并在系统电源和IIDG接入处安

装有电压互感器获得电压的有效值,这些测量值也能反映配电网的运行方式。考虑配电网实际测量条件,分别采用上述2类电气量作为配电网运行方式特征信息,再结合故障位置及过渡电阻,构成样本的输入特征。故障类型则不作为输入特征,因为不同故障类型下配电网电气量特征差异较大,样本混合训练会降低模型性能,为此,与传统短路电流计算习惯一致,针对不同故障类型调用不同的计算模型。样本的标签为对应上述输入特征的各项线路短路电流稳态值。

综上所述,对于包含 K 个节点、 E 条支路的配电网,假设有1个系统电源接入节点和 H 个IIDG接入节点,则有2类短路电流计算模型输入特征,即 X_1 对应配置 μ PMU的配电网, X_2 对应未配置 μ PMU的配电网,分别如式(8)和式(9)所示。

$$X_1 = \left\{ \left| V_1 \right|, \dots, \left| V_k \right|, \dots, \left| V_K \right|, \theta_1, \dots, \theta_k, \dots, \theta_K, \right. \\ \left. P_1, \dots, P_e, \dots, P_E, Q_1, \dots, Q_e, \dots, Q_E, \right. \\ \left. F_{\text{line}}, F_{\text{loc}}, R_{\text{on}}, S_{\text{DG}_1}, \dots, S_{\text{DG}_h}, \dots, S_{\text{DG}_H} \right\} \quad (8)$$

$$X_2 = \left\{ I_1, \dots, I_e, \dots, I_E, V_1, V_{\text{DG}_1}, \dots, V_{\text{DG}_h}, \dots, V_{\text{DG}_H}, \right. \\ \left. F_{\text{line}}, F_{\text{loc}}, R_{\text{on}}, S_{\text{DG}_1}, \dots, S_{\text{DG}_h}, \dots, S_{\text{DG}_H} \right\} \quad (9)$$

式中:下标 k 为节点编号,节点1为系统电源接入节点;下标 e 为支路编号; $|V|$ 、 θ 、 P 、 Q 、 I 、 V 分别包含三相电压幅值、电压相位、有功、无功、支路电流、电压有效值数据,因此能应用于不对称配电网; F_{line} 为故障线路编号; F_{loc} 为故障位置距离线路首端的位置百分比, $F_{\text{loc}} \in [0, 100\%)$,当 $F_{\text{loc}}=0$ 时,认为故障发生在首节点上; R_{on} 为过渡电阻,发生接地故障时为相与地之间的电阻,仅相间故障时为相与相之间的电阻; S_{DG_h} 为第 h 个IIDG的容量。

采用基于CNN的算法适应方法时,为了更好地发挥CNN的局部敏感特性,将输入数据特征由一维修改至二维。具体是将反映运行方式的特征数据分为不同数据层,式(8)中分为了电压幅值层、电压相位层、有功层和无功层,这些层以节点或支路编号对齐,且每层有3行分别对应三相。故障信息与IIDG信息单独为1层,该层为稀疏矩阵,仅有故障和IIDG对应的节点或支路编号有相应信息,其余均为0。式(9)类似,只是将反映运行方式的数据分为了电流层、电压层。

2.2 评价指标

计算准确性一般由真实值 y_j 与计算值 $\hat{y}_j$ 的差异来衡量,其中真实值通常由仿真获得。通常采用平均绝对误差MAE(Mean Absolute Error)、平均绝对百分比误差MAPE(Mean Absolute Percentage Error)来衡量模型准确度。MAE和MAPE指标都受真实值影响:当真实值大时,MAE受影响大;当真实

值小时,MAPE受影响大。因此用MAE和MAPE这两个指标综合反映算法性能,避免真实值过大或过小带来的影响。

多输出模型中第 j 个输出的MAE值 δ_j 定义为:

$$\delta_j = \frac{1}{N} \sum_{i=1}^N |(\hat{y}_j^i - y_j^i)| \quad j \in \{1, 2, \dots, D\} \quad (10)$$

MAPE是由绝对百分比误差APE(Absolute Percentage Error)演化而来,相关定义为:

$$\lambda_{APE} = \left| \frac{\hat{y}_j^i - y_j^i}{y_j^i} \right| \times 100\% \quad (11)$$

$$\lambda_j = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_j^i - y_j^i}{y_j^i} \right| \times 100\% \quad (12)$$

式中: λ_{APE} 为APE值; λ_j 为第 j 个输出的MAPE。

上述评价指标的数值越小,模型准确性越高。多输出模型会产生 D 个 δ_j 和 λ_j ,为方便展示,文中选用 D 个 δ_j 、 λ_j 指标中的最大值 δ_{MAX} 、 λ_{MAX} 和平均值 δ_{MEAN} 、 λ_{MEAN} 。

2.3 模型参数寻优

对于机器学习而言,超参数是模型训练过程中需要重点调整的内容,获得最优的超参数可以提升模型性能,使评价指标 δ_{MEAN} 和 λ_{MEAN} 达到最小。不同类型的多输出模型采用的超参数寻优方式不完全相同。

对于问题转化方法,本文采用随机采样网格搜索与5折交叉验证结合的方式进行超参数寻优,以最终输出的 λ_{MEAN} 作为评价标准,验证超参数寻优结果。XGBoost最佳超参数组合为:基分类器为400个,树最大深度为10,子节点权重为2,子采样系数为0.8,特征采样系数为0.8。LightGBM最佳超参数组合为:基分类器为430个,叶子数为20,子采样系数为0.8,特征采样系数为0.9。

对于基于神经网络的多输出模型,超参数更为复杂,为避免网格搜索陷入局部最优,调参选用效率更高的贝叶斯调参法。本文使用3层全连接的ANN,每一层都需要使用激活函数,激活函数选择ReLU函数,学习率为 10^{-5} ,单次训练样本100个。CNN最优参数组合为:卷积核大小为 3×3 ,填充半径为1,激活函数为ReLU,池化层大小为2,学习率为 10^{-5} ,单次训练样本150个。CNN输出层需要将前层卷积输出展开,使用1层全连接层连接,采用ReLU激活获得最终结果。基于神经网络的多输出模型均使用 λ_{MEAN} 作为损失函数。

2.4 多输出模型计算流程

含IIDG配电网的短路电流多输出计算模型包括训练和应用两部分。其中,训练需要大量样本,通过仿真获得样本的方法同文献[11],有所不同的是,提取的输入特征有差异,且该文献的样本标签为1个测量点,而本文的样本标签为所有支路测量点。

具体流程如下:

1)通过MATLAB/Simulink建立所需的配电网模型;

2)在常见运行方式区间内,随机设置配电网运行方式和故障信息;

3)进行故障仿真,并记录输入特征和各条支路的短路电流稳态值作为标签值。

重复上述步骤2)、3)即可获取1组训练样本。训练样本的数量可以根据训练效果进行补充或调整。为了减少不同输入特征数值大小对模型训练的影响,对输入数据进行归一化预处理。模型训练过程本质是调参过程,寻找模型最优参数使得输出结果误差最优,达到预期则训练结束。模型应用时,输入配电网的运行方式特征信息、故障信息,根据故障类型调用训练好的对应模型,则可输出所有支路的短路电流值。

3 算例验证

3.1 算例情况

含IIDG配电网结构图如图1所示,按照图1以IEEE 34节点配电网系统为基础接入IIDG,IIDG容量设定为: $S_{DG_1}=500$ kW, $S_{DG_2}=200$ kW, $S_{DG_3}=500$ kW, $S_{DG_4}=400$ kW。节点800为系统电源接入点,考虑最大和最小2种基础运行方式,等值阻抗分别为 $j0.5 \Omega$ 和 $j1 \Omega$ 。配电网电压等级为24.9 kV,系统线路共有31条,负荷参照文献[16]设定,在最大、最小运行方式下不发生改变。

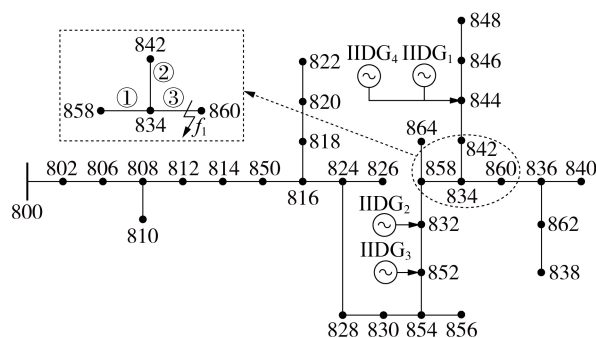


图1 含IIDG配电网结构图

Fig.1 Structure of distribution network with IIDGs

随机设置配电网运行方式,模型参数设置同文献[11]:等值电源阻抗在系统最大、最小等值阻抗区间内随机产生;IIDG容量及负荷大小则在其原值的 $[80\%, 120\%]$ 范围内随机产生;设置一定概率的随机切除部分线路或IIDG。随机设置故障线路编号 F_{line} 、故障发生位置 F_{loc} 和过渡电阻 R_{on} 。上述运行方式和故障信息的设置,均会通过式(8)或式(9)的输入特征反映出来。需说明的是,过渡电阻的设置范围应与故障类型相吻合,例如配电网单相接地短路

故障的过渡电阻范围可设置为 $[0, 1000] \Omega^{[17]}$,两相相间短路故障的过渡电阻可设置更小的取值范围,三相短路故障可不设置。

本算例以三相短路电流为例,通过仿真获得10000组样本。由于仿真建模时IIDG采用较为精确的模型,因此仿真速度很慢,获取10000组样本耗时约10 d。按照8:2的比例将样本分为训练集和测试集。根据式(8),该配电网共有316个输入特征,根据式(9),则有91个输入特征,输出结果为80个。

3.2 多输出模型有效性验证

文献[11]采用XGBoost和LightGBM模型进行预测,并已与物理建模方法进行了对比,结果表明机器学习计算短路电流的准确性。但文献[11]为单输出模型,若要输出图1虚线框中测量点①、②、③的短路电流,则需要建立3个模型,而本文所提方法仅需1个模型即可输出上述3处测量点的短路电流。表1为不同方法预测节点834发生三相短路故障时C相短路电流的结果与模型的性能指标。其中,XGBoost(单输出)模型的数据来源于文献[11],为测量点①的电流,MTRS-XGBoost(多输出)和MTRS-LightGBM(多输出)模型的结果为测量点①、②、③处的预测结果。因本文与文献[11]的系统等值阻抗、负荷、IIDG容量等均在相同范围内随机生成,故而运行方式差异不会太大但无法保证完全一致,因此选择了实际值较为接近的场景来进行性能比较。

表1 单输出模型与多输出模型结果

Table 1 Result of single-output model and multi-output model

模型	测量点	实际值 / A	预测值 / A	$\lambda_{APE} / \%$
XGBoost(单输出)	①	240.85	241.72	0.364
MTRS-XGBoost (多输出)	①	242.06	241.89	0.070
	②	18.33	18.36	0.136
	③	238.43	238.34	0.038
MTRS-LightGBM (多输出)	①	242.06	241.82	0.097
	②	18.33	18.32	0.043
	③	238.43	238.28	0.065

表2 不同多输出模型、不同基学习器、不同输入特征的性能对比

Table 2 Performance comparison of different multi-output models, base learners and input characteristics

多输出模型	基学习器	输入特征	$\lambda_{MEAN} / \%$	$\lambda_{MAX} / \%$	δ_{MEAN} / A	δ_{MAX} / A	训练时间 / s	测试时间 / s
问题转化-MTRS	XGBoost	式(8)	0.001 25	0.019 07	0.010 24	0.023 27	661.977	9.379 31
		式(9)	0.003 15	0.116 15	0.010 71	0.023 87	254.683	11.750 62
	LightGBM	式(8)	0.001 69	0.030 23	0.012 65	0.028 16	1 027.101	8.713 61
		式(9)	0.003 07	0.130 04	0.012 49	0.026 99	270.293	10.010 61
问题转化-RC	XGBoost	式(8)	0.001 46	0.022 84	0.012 87	0.022 98	1 279.650	0.602 41
		式(9)	0.002 31	0.072 43	0.012 93	0.024 65	538.382	0.454 81
	LightGBM	式(8)	0.001 89	0.023 93	0.015 46	0.029 22	917.070	3.883 91
		式(9)	0.002 89	0.121 07	0.014 75	0.027 85	414.073	4.338 82
神经网络 多输出	ANN	式(8)	0.092 59	0.853 61	0.075 67	0.148 70	3 503.504	0.064 82
		式(9)	0.103 39	0.968 59	0.156 47	0.231 96	2 847.905	0.054 89
	CNN	式(8)	0.186 19	1.610 94	0.513 26	1.273 76	982.345	0.028 48
		式(9)	0.192 27	1.632 82	0.355 84	0.947 37	461.215	0.016 92

由表1可知,XGBoost(单输出)模型只能预测单个测量点的短路电流, λ_{APE} 在可接受范围内;而多输出模型预测的短路电流,因考虑各输出间的相关性, λ_{APE} 较单输出模型更低。以测量点①为例,MTRS-XGBoost(多输出)和MTRS-LightGBM(多输出)模型比XGBoost(单输出)模型的 λ_{APE} 小,分别降低了80.7%和73.4%。此外,对于不同测量点的短路电流,使用不同基学习器模型时短路电流误差情况略有不同。综上所述,采用多输出模型计算含IIDG配电网短路电流不仅有效,且准确性更高。

此外,某运行方式下测量点①、②、③的三相短路电流相量如附录A图A1所示。由图可见,IIDG的接入使得各支路稳态电流的幅值不再显式地满足基尔霍夫电流定律,但相量依然满足基尔霍夫电流定律。

3.3 多输出模型性能对比

下面对基于问题转化方法和算法适应方法的2类多输出模型的性能进行对比,同时对比式(8)和式(9)所示的输入特征对计算性能的影响,结果如表2所示。表中:训练时间为完成模型训练的时间;测试时间为完成2000组测试的总耗时;加粗数值表示每种多输出模型下的性能最优项。对表2进行分析可得到下列结论。

1)6种模型和2种输入特征均取得了较好的性能; δ_{MAX} 、 δ_{MEAN} 、 λ_{MAX} 和 λ_{MEAN} 均在可接受范围内;最长测试时间为11.750 62 s,即最长平均测试时间约为5.875 ms,而单输出模型的平均预测时间为0.049 s^[11],因此多输出模型预测速度快,约为单输出模型的8.4倍。

2)对比同一种多输出模型使用不同输入特征的计算性能可知,6种模型选用式(8)所示的输入特征时,其MAPE指标优于选用式(9)所示的输入特征时的MAPE指标;采用问题转化方法时,多输出模型的MAE比较接近,而采用基于神经网络的算法适应方法时,多输出模型的MAE有一定差异,且选用式

(8)、(9)所示的输入特征均有可能获得相对较优的结果。因此,2类输入特征对多输出模型的准确度影响不大,在尚未普及 μ PMU 配置时,可以使用式(9)所示的输入特征。

3)在问题转化方法中,2种基学习器的性能在准确度方面差异并不大,但是在速度却有快慢之分。采用 MTRS 方法时,在相同的输入特征下,LightGBM 的速度快于 XGBoost,这证明在输入特征变动不频繁的情况下,LightGBM 速度更快,但也因为 EFB 特性导致与 XGBoost 的速度并没有明显的区分。采用 RC 方法时,EFB 特性反而成为了 LightGBM 提升速度的障碍。

4)对比同一种输入特征下不同多输出模型的性能可以发现:2类问题转化方法中,XGBoost 基学习器的误差略小于 LightGBM 基学习器,计算速度方面更多地取决于所采用的问题转化方法,应用时 RC 方法明显快于 MTRS 方法;算法适应方法中,ANN 比 CNN 有更低的误差,而计算速度不如 CNN,上述结论与理论分析一致。算法适应方法的准确度不如问题转化方法,但应用速度快于问题转化方法。综上,针对含 IIDG 配电网短路电流计算问题,采用问题转化方法的准确性都非常高,计算速度最快的是选用 XGBoost 基学习器的 RC 方法。而神经网络多输出方法适用于对准确性要求一般,而对应用速度要求极高的场合。此外,选取测量点①、②、③处的 A 相电流的真实值与不同算法输出进行对比,结果见附录 A 图 A2。由图可见,故障线路上基于 ANN 和 CNN 的短路电流预测值偏差较为明显,其他差异不太明显。

3.4 模型抗干扰能力

在实际应用时,数据采集过程可能出现部分特征数据丢失的情况。选用表 2 中综合性能较优的方法,即分别使用 MTRS、RC 拼合的 XGBoost 算法和 ANN 多输出算法,输入特征选用式(8),数据丢失以随机置 0 的方式表示。3 种模型在不同数据丢失比例 r_{loss} 下的性能表现如表 3 和附录 A 图 A3 所示,表 3 中加粗数值为较小误差数据。

表 3 表明:在丢失 1% 的数据量时,3 种模型的误差均较小,当丢失数据量大于等于 3% 时,ANN 多输出模型的误差增大,准确性明显降低;使用问题转化的多输出模型在丢失 5%、10% 的数据量情况下仍能保持较优的性能,其中 MTRS 模型误差略优于 RC 模型。

图 A3 为数据丢失比例高于 10% 的情况下,不同多输出模型的 λ_{MEAN} 。由图可见:当数据丢失比例在 30% 以下时,MTRS-XGBoost 模型和 RC-XGBoost 模型的 λ_{MEAN} 均较小;当数据丢失比例超过 30% 后,MTRS-XGBoost 模型和 RC-XGBoost 模型的 λ_{MEAN} 快

表 3 不同数据丢失比例下多输出模型的性能对比

Table 3 Performance comparison among multi-output models with different rates of data loss

$r_{\text{loss}} / \%$	性能指标	指标值		
		MTRS-XGBoost	RC-XGBoost	ANN 多输出
1	λ_{MAX}	0.260	0.272	0.810
	λ_{MEAN}	0.036	0.041	0.349
	δ_{MAX}	0.053	0.054	0.592
	δ_{MEAN}	0.031	0.038	0.441
3	λ_{MAX}	0.624	0.644	2.061
	λ_{MEAN}	0.113	0.121	1.582
	δ_{MAX}	0.098	0.101	2.736
	δ_{MEAN}	0.063	0.078	2.054
5	λ_{MAX}	0.898	0.885	3.621
	λ_{MEAN}	0.163	0.194	3.102
	δ_{MAX}	0.151	0.154	5.360
	δ_{MEAN}	0.091	0.109	4.021
10	λ_{MAX}	1.859	2.045	5.899
	λ_{MEAN}	0.332	0.400	5.818
	δ_{MAX}	0.268	0.300	10.517
	δ_{MEAN}	0.180	0.199	7.902

速上升;当数据丢失比例为 60% 时,MTRS-XGBoost 模型和 RC-XGBoost 模型 λ_{MEAN} 与 ANN 多输出模型相近。由此可知,当数据少量丢失的情况下,MTRS 和 RC 方法均具有较好的抗干扰能力,当数据丢失比例较高时,2 种方法的误差随数据丢失比例的升高而快速增大。

4 结论

含 IIDG 配电网发生短路故障时,IIDG 的输出具有较强非线性特征,使得短路电流的迭代计算耗时增长,而采用简化模型又影响计算的准确性。在 IIDG 高占比接入配电网的发展趋势下,本文提出了数据驱动的含 IIDG 配电网短路电流计算多输出模型,解决了计算准确性与计算速度之间的矛盾,得到如下结论。

1)针对含 IIDG 配电网短路电流计算,选用多输出模型,不仅解决了单输出模型的模型数量过多的问题,且具有更高的准确性和更快的计算速度,例如采用 MTRS-XGBoost 模型时 λ_{APE} 降低了 80.7%,最长测试时间下也比单输出模型快 8.4 倍,因此本文所提出多输出模型能满足在线应用要求。

2)无论在常规的配电网数据采集条件,还是配置有 μ PMU 的条件下,都可以满足所提多输出模型的输入特征条件;即使在输入数据丢失 30% 的情况下, λ_{MEAN} 仍能达到小于 5% 的要求。

3)相比较而言,基于 MTRS 方法的多输出模型具有准确性高和抗干扰能力强的优点,虽然计算速度不及其他多输出模型,但也远快于单输出模型和物理建模计算速度。RC 方法的准确性和抗干扰能力与 MTRS 方法相近,但计算速度较 MTRS 方法快

1个数量级,是较为稳定的多输出方法。使用神经网络的算法适应方法具有最快的应用速度,但是其准确度相对略低,适用于对计算快速性要求极高的场景。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 王子欣,苗世洪,郭舒毓,等.考虑分布式电源出力随机特性的配电网节点脆弱性评估[J].电力自动化设备,2021,41(8):33-40.
WANG Zixin, MIAO Shihong, GUO Shuyun, et al. Node vulnerability evaluation of distribution network considering randomness characteristic of distributed generation output [J]. Electric Power Automation Equipment, 2021, 41(8): 33-40.
- [2] JIA Ke, YANG Zhe, FANG Yu, et al. Influence of inverter-interfaced renewable energy generators on directional relay and an improved scheme[J]. IEEE Transactions on Power Electronics, 2019, 34(12): 11843-11855.
- [3] 习伟,张哲,陈卫,等.光伏逆变器短路特性分析[J].电力系统保护与控制,2015,43(7):70-75.
XI Wei, ZHANG Zhe, CHEN Wei, et al. Research on fault characteristic of the PV inverter[J]. Power System Protection and Control, 2015, 43(7): 70-75.
- [4] 解腾,徐晔,王金全,等.脉冲负载下逆变型分布式电源故障电流特性分析[J].科学技术与工程,2018,18(27):147-151.
XIE Teng, XU Ye, WANG Jinquan, et al. Analysis of fault current characteristics of inverter interfaced distributed generation applying to pulse load[J]. Science Technology and Engineering, 2018, 18(27): 147-151.
- [5] 魏传芝,施啸寒,张恒旭,等.考虑电流指令跟踪能力的PQ控制逆变型分布式电源故障模型[J].电力自动化设备,2020,40(1):59-65,72.
WEI Chuazhi, SHI Xiaohan, ZHANG Hengxu, et al. Fault model of IIDG with PQ control considering tracking capability of current command[J]. Electric Power Automation Equipment, 2020, 40(1): 59-65, 72.
- [6] 常仲学,杨忠礼,宋国兵,等.正负双序独立控制策略下的逆变型分布式电源不对称故障电流分析[J].电力自动化设备,2018,38(1):44-51.
CHANG Zhongxue, YANG Zhongli, SONG Guobing, et al. Analysis on asymmetric fault current characteristics of inverter interfaced distributed generator under positive-sequence and negative-sequence current respective control strategy[J]. Electric Power Automation Equipment, 2018, 38(1): 44-51.
- [7] 孔祥平,张哲,尹项根,等.含逆变型分布式电源的电网故障电流特性与故障分析方法研究[J].中国电机工程学报,2013,33(34):65-74, S9.
KONG Xiangping, ZHANG Zhe, YIN Xianggen, et al. Study on fault current characteristics and fault analysis method of power grid with inverter interfaced distributed generation[J]. Proceedings of the CSEE, 2013, 33(34): 65-74, S9.
- [8] 谭会征,李永丽,陈晓龙,等.带低电压穿越特性的逆变型分布式电源对配电网短路电流的影响[J].电力自动化设备,2015,35(8):31-37, 52.
TAN Huizheng, LI Yongli, CHEN Xiaolong, et al. Influence of inverter-interfaced distributed generator with low-voltage ride-through capability on short circuit current of distribution network[J]. Electric Power Automation Equipment, 2015, 35(8): 31-37, 52.
- [9] 杨杉,同向前.含低电压穿越型分布式电源配电网的短路电流计算方法[J].电力系统自动化,2016,40(11):93-99.
YANG Shan, TONG Xiangqian. Short-circuit current calculation of distribution network containing distributed generators with capability of low voltage ride through[J]. Automation of Electric Power Systems, 2016, 40(11): 93-99.
- [10] 金基达,宋依群,范春菊,等.考虑逆变电源控制策略的电流保护整定计算[J].电网技术,2021,45(9):3690-3699.
JIN Shenda, SONG Yiqun, FAN Chunju, et al. Calculation of current protection setting based on inverter generation control strategy[J]. Power System Technology, 2021, 45(9): 3690-3699.
- [11] 郑翔,王慧芳,姜宽,等.机理与数据融合驱动的含IIDG配电网短路电流计算方法[J].电力自动化设备,2021,41(1):41-48.
ZHENG Xiang, WANG Huifang, JIANG Kuan, et al. Mechanism and data-driven combined short circuit current calculation method for distribution network with IIDG[J]. Electric Power Automation Equipment, 2021, 41(1): 41-48.
- [12] ZHENG Xiang, WANG Huifang, JIANG Kuan, et al. A novel machine learning-based short-circuit current prediction method for active distribution networks [J]. Energies, 2019, 12(19): 3793.
- [13] BORCHANI H, VARANDO G, BIELZA C, et al. A survey on multi-output regression[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2015, 5(5): 216-233.
- [14] 张又文,冯斌,陈页,等.基于遗传算法优化XGBoost的油浸式变压器故障诊断方法[J].电力自动化设备,2021,41(2):200-206.
ZHANG Youwen, FENG Bin, CHEN Ye, et al. Fault diagnosis method for oil-immersed transformer based on XGBoost optimized by genetic algorithm[J]. Electric Power Automation Equipment, 2021, 41(2): 200-206.
- [15] 郑一斌,王慧芳,张磊,等.基于LightGBM算法的配电网单相接地故障区段定位方法[J].电力自动化设备,2021,41(12):54-61.
ZHENG Yibin, WANG Huifang, ZHANG Lei, et al. Single-phase grounding fault section location in distribution network based on LightGBM algorithm[J]. Electric Power Automation Equipment, 2021, 41(12): 54-61.
- [16] IEEE Power Energy Systems Distribution System Analysis Subcommittee. Distribution test feeder[EB/OL]. (2010-09-17) [2021-02-05]. <http://www.ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html>.
- [17] 熊卫红,毛兴华,李景禄,等.小电阻接地方式对人身安全的影响及智能电阻接地方式研究[J].电力系统保护与控制,2019,47(14):166-172.
XIONG Weihong, MAO Xinghua, LI Jinglu, et al. Influence of small resistance grounding mode on personal safety and research of intelligent resistance grounding mode[J]. Power System Protection and Control, 2019, 47(14): 166-172.

作者简介:



叶睿恺

叶睿恺(1992—),男,博士研究生,主要研究方向为电力系统数据挖掘与人工智能应用(E-mail: pokagic@zju.edu.cn);

王慧芳(1974—),女,副教授,博士,通信作者,主要研究方向为电网状态检修、继电保护与控制、电力系统数据挖掘(E-mail: huifangwang@zju.edu.cn);

张森(1998—),男,硕士研究生,主要研究方向为智能电网数据挖掘、人工智能应用(E-mail: zzs98@zju.edu.cn)。

(编辑 任思思)

(下转第132页 continued on page 132)

algorithm based analysis on massive data of intelligent power utilization[J]. Power System Technology, 2014, 38(10): 2715-2720.

- [20] 尚学军, 霍现旭, 郑晓冬, 等. 基于离散小波分析与K-means聚类算法的MMC-HVDC输电线路保护方案[J]. 电测与仪表, 2020, 57(24): 52-57.

SHANG Xuejun, HUO Xianxu, ZHENG Xiaodong, et al. MMC-HVDC transmission line protection scheme based on discrete wavelet analysis and K-means clustering algorithm[J]. Electrical Measurement & Instrumentation, 2020, 57(24): 52-57.

作者简介:



戴志辉

戴志辉(1980—),男,教授,博士研究生导师,博士,主要研究方向为电力系统保护与控制(E-mail: daihuadian@163.com);

陈思琦(1994—),女,硕士研究生,主要研究方向为电力系统保护与控制(E-mail: 13785381821@163.com)。

(编辑 任思思)

Protection principle of flexible DC distribution line based on clustering center distance measurement

DAI Zhihui, CHEN Siqi, LI Yiran, ZHANG Yihong, HE Jingyuan, YU Lirui

(Hebei Key Laboratory of Distributed Energy Storage and Microgrid, North China Electric Power University, Baoding 071003, China)

Abstract: Aiming at the issues of poor selectivity of single-terminal protection, and poor speed and high requirements for data synchronization of double-terminal protection in existing flexible DC distribution network, the protection principle of flexible DC distribution line based on clustering center distance measurement is proposed. The principle only uses single-side current as the characteristic quantity. Different types of short circuit faults at different locations of the line are simulated, and combined with the actual faults, the post-fault line currents are collected and processed to form historical data samples. The K-means clustering algorithm is then used to find the best clustering centers under different types of faults, and fault identification and pole selection are realized by comparing the distance between real-time data and each clustering center. The proposed principle doesn't need complex feature extraction and calculation process, and the complex setting of conventional current protection can be avoided. Finally, the simulation is carried out in PSCAD/EMTDC, and the results show that the proposed protection principle has good selectivity and rapidity, and the ability to resist the transition resistance is improved compared with the conventional current protection based on local information.

Key words: flexible DC distribution network; clustering algorithms; clustering center; relay protection; single-terminal protection; fault identification

(上接第125页 continued from page 125)

Data-driven multi-output model for short circuit current calculation in distribution network with IIDGs

YE Ruikai, WANG Huifang, ZHANG Sen, ZHANG Yixiang

(College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China)

Abstract: A large number of IIDGs (Inverter Interfaced Distributed Generators) are connected to the distribution network under the "dual-carbon" target, which leads to an increasingly serious contradiction between the calculation speed and accuracy in short circuit current calculation of distribution network based on mechanism modeling method. Data-driven modeling method has been proved to be able to solve the contradiction effectively. However, single-output model is difficult to meet the demand of multiple calculation output in practical applications, which may cause a problem of model quantity. Aiming at the above problem, the multi-output regression modeling and calculation method for short circuit current in distribution network is proposed. The problem transformation method and algorithm adaptation method are analyzed and compared, the MTRS (Multi-Target Regressor Stacking), RC (Regressor Chains) and other problem transformation methods and multi-output neural network method are proposed, which are suitable for solving the short circuit current calculation problem of distribution network with IIDGs. The pivotal issues such as input feature selection, performance evaluation index, calculation process and optimization method of hyper parameters are analyzed. The example shows that the multi-output modeling method can satisfy calculation speed and accuracy simultaneously, and the performance of multi-output model is better than that of single-output model, and the problem of model quantity in single-output modeling method is avoided.

Key words: short circuit current calculation; multi-output regression; inverter interfaced distributed generator; distribution network; data-driven modeling

附录 A

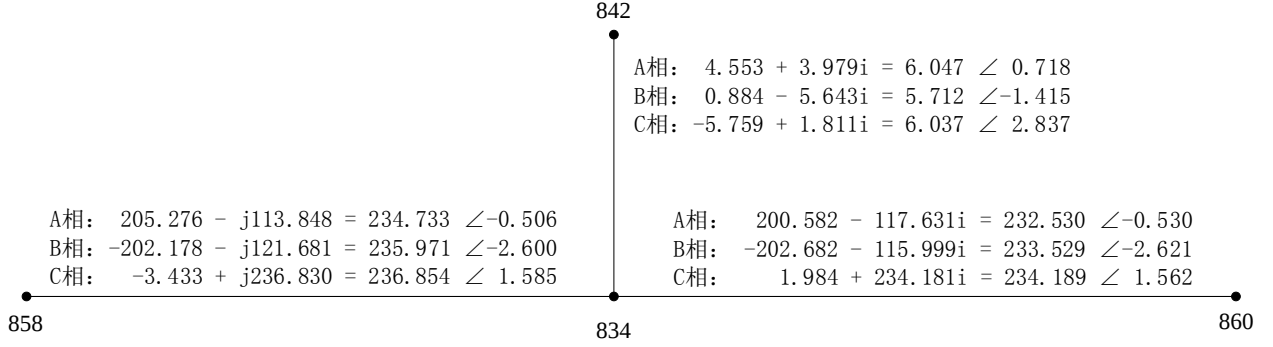


图 A1 含 IIDG 配电网短路电流分布示例

Fig.A1 Short circuit current distribution diagram of distribution network with IIDGs

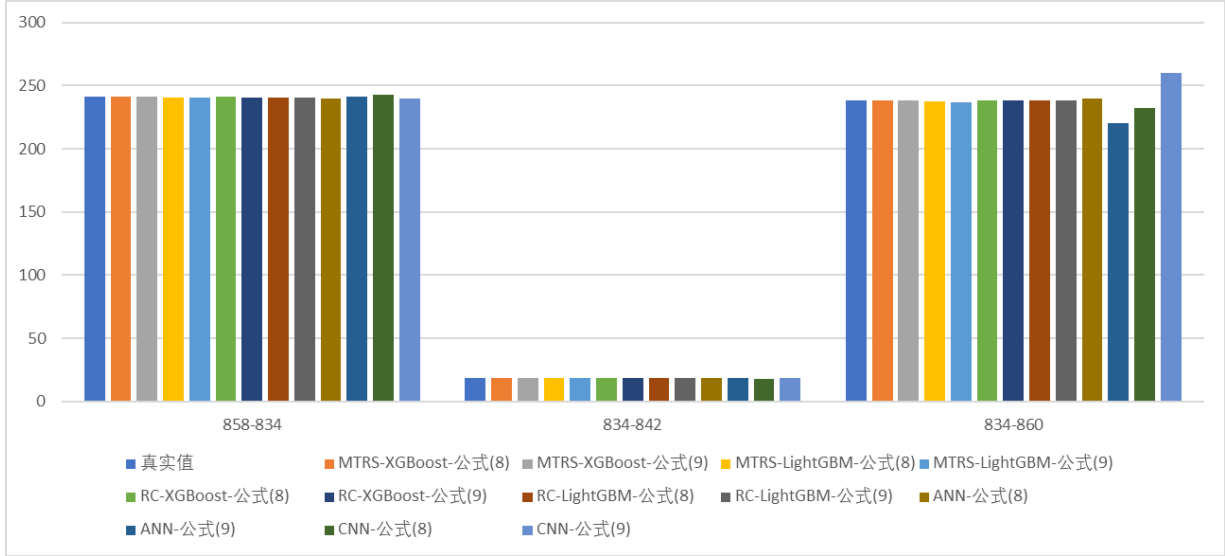


图 A2 不同算法的预测值与真实值对比图

Fig.A2 Performance comparison between prediction value of different algorithms and real values

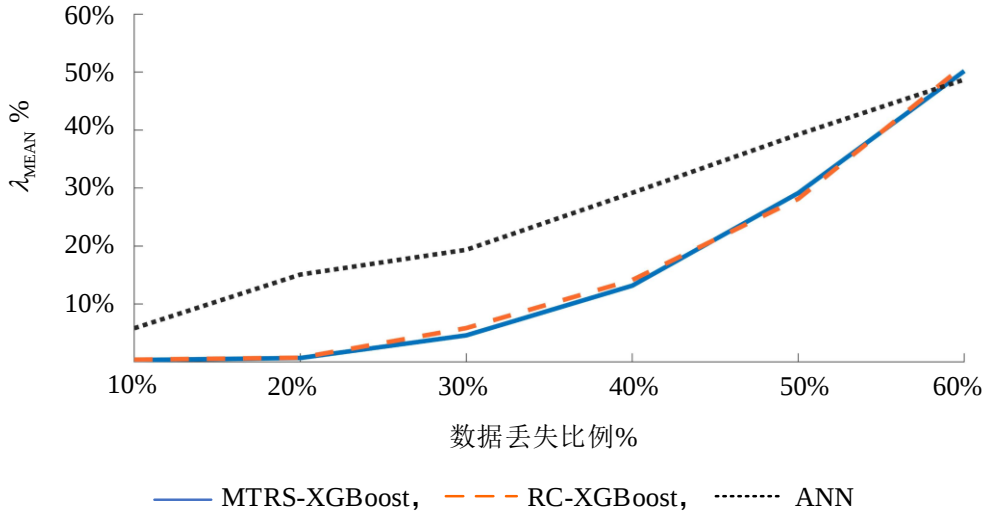


图 A3 高数据丢失比例下多输出模型的性能对比

Fig.A3 Comparison of multi-output models under high proportion of data losing