机理与数据融合驱动的含 IIDG 配电网短路电流计算方法

郑 翔¹,王慧芳¹,姜 宽¹,曹建伟²,刘 莹²,陈永炜² (1. 浙江大学 电气工程学院,浙江 杭州 310027;2. 国网湖州供电公司,浙江 湖州 313000)

摘要:目前含逆变器型分布式电源(IIDG)的配电网短路电流计算主要采用物理建模方法,在IIDG高渗透情况下其在计算速度、准确性、通用性等方面存在不足。因此提出一种机理与数据融合驱动的适用于含IIDG和IIDG高渗透下配电网的短路电流计算方法。为与数据驱动建模的计算方法进行对比,在对含IIDG的配电网进行特征分析的基础上,提出了反映短路电流的2种特征组合方式:一种使用蕴含机理的特征,即将不接入IIDG时配电网的短路电流作为关键特征,另一种则使用配电网特征。通过运行MATLAB/Simulink上搭建的仿真模型自动积累样本集合,使用机器学习中的集成方法(包括随机森林、极限随机树、XGBoost、LightGBM)进行2种特征下的模型训练。在IEEE 34节点系统上验证了集成方法建模的可行性和有效性,同时对比了不同集成方法以及不同特征组合方式的计算误差,结果表明,各集成学习方法均能够准确地进行短路电流计算,机理与数据融合的驱动方法在机理未失效情况下,比单纯的数据驱动模型计算更准确。与物理建模方法的对比结果也验证了所提方法的准确性和快速性。

0 引言

越来越多的分布式电源通过电力电子逆变器并 入配电网,这类电源被称为逆变器型分布式电源IIDG (Inverter-Interfaced Distributed Generator)。IIDG具 有强非线性特征,故配电网发生短路时,其输出的短 路电流与其他电源区别明显^[1]。目前配电网中IIDG 渗透率不高,传统短路电流计算方法误差不大,但随 着 IIDG 渗透率的不断提高,传统方法的误差将难以 满足应用要求。因此,研究 IIDG 高渗透的配电网的 短路电流计算具有重要的理论和应用意义。

对于含 IIDG 配电网短路电流计算问题,国内外 学者展开了大量的研究:文献[2]将 IIDG 假设为故 障后恒功率运行的电源,通过潮流计算获得 IIDG 在 短路时不能突变的状态量来求取 IIDG 向电网提供 的电流,但该文献没有考虑 IIDG 的低电压穿越 (LVRT)特性和限流作用;文献[3]将基于PQ 控制的 IIDG 等效为压控电流源模型,建立节点电压方程, 采用迭代修正求解算法进行求解;文献[4]将 IIDG 建立为提供 LVRT 无功支持的电流源模型,并考虑 各 IIDG 脱网时间不同对短路电流的影响,采用三序 网络叠加的迭代算法;文献[5]考虑 IIDG 的 LVRT特 性,将 IIDG 分为故障点上游和故障点下游分别建立 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。以上研究主要使用物理建模的方法,首

收稿日期:2020-02-08;修回日期:2020-08-20

基金项目:国网湖州供电公司科技项目(2019-HUZJTKJ-17) Project supported by the Science and Technology Program of State Grid Huzhou Power Supply Company(2019-HUZJTKJ-17) 先建立 IIDG 在故障发生时的等效模型,进而求解短路电流。但由于 IIDG 等效模型十分复杂,往往会根据应用场景对模型进行一定的简化,因此模型的适用性和准确性受到限制。同时,由于 IIDG 等效模型的输出电流受到并网点电压等因素的影响,在求解短路电流时需要使用迭代算法不断修正直到满足精度要求,因此计算效率较低。随着 IIDG 渗透率的提高及控制策略的日趋复杂,物理建模方法在计算速度、准确性、通用性方面的不足愈来愈明显。

近年来,人工智能,尤其是机器学习方法得到了 越来越多的关注和应用[6-7]。机器学习是一种数据 驱动的建模方法,可以根据已知数据,分析得到输入 与输出之间的依赖关系,继而对未知数据进行预测。 因此机器学习可以脱离物理建模,直接从数据中挖 掘信息进而得到结果。目前在难以通过机理分析建 模的复杂问题中,机器学习已经有较为出色的表 现^[8-11],这给含 IIDG 和 IIDG 高渗透的配电网短路电 流快速计算问题带来了新的思路。对于特定配电网 的某种稳态运行方式,当故障信息如故障位置、故障 类型等确定时,各个点流过的短路电流是唯一的,数 值可以通过仿真获得,因此可以使用机器学习方法 自动学习配电网特征与短路电流之间的映射关系。 此外,数据驱动建模能够把准确计算故障电流消耗 的时间集中在离线进行的样本获得和模型训练上, 在线应用时可以非常快速地获得准确的结果。

然而,目前基于机器学习的数据驱动建模在电 力领域仅用于负荷预测、光伏预测等相对新的领 域,还未应用于短路电流等传统计算领域,再加上 该方法存在可解释性不足的缺点,因此不容易被传 统领域的专家接受。为此,本文提出机理与数据融合驱动的含 IIDG 配电网的短路电流计算方法。首先分析与短路电流相关的配电网特征,提出将不接入 IIDG 时配电网的短路电流作为关键机理特征,为进行对比,构建了2种特征组合。然后,利用 MATLAB / Simulink进行电网建模及仿真计算,自动形成训练样本集。接着,在分析常用的机器学习方法基础上,采用集成方法(包括随机森林、极限随机树、极端梯度提升(XGBoost)、LightGBM)离线训练2 类不同特征的机器学习模型,进行在线应用。最后,利用算例验证了机理与数据融合驱动的含 IIDG 配电网的短路电流计算模型的有效性。

1 特征分析与样本组成

在进行机器学习前,首先需要确定组成样本的 特征和标签。本应用中,流经短路电流测量点的短 路电流即为样本标签,测量点可由用户设定,如某继 电保护关心的电流。特征是影响短路电流结果大小 的配电网特征,通常包含配电网稳态特征和故障特 征。配电网稳态特征既包括反映系统稳态运行状态 的电气量,如节点电压幅值和相角、线路传输的功 率、旋转型电源的等值阻抗、负荷功率、IIDG容量及 注入电流等,也包括配电网拓扑结构特征,如IIDG 投入运行情况、线路投切情况等。故障特征则是故障 的相关信息,包括故障位置、故障类型、过渡电阻等。 由于进行短路电流计算时,通常将故障位置、故障类 型等信息作为已知条件,因此可不考虑故障特征。

上述配电网特征虽然比较全面,但是数量较多, 且会随着配电网规模的扩大而增加,当配电网十分 复杂时,特征数量将非常巨大,不利于机器学习。另 一方面,这些特征中往往只有一小部分特征会对短 路电流产生较大影响,大部分特征存在冗余,质量较 低,将降低机器学习模型的性能。因此,笔者希望能 够找到数量更少、质量更高的特征组合。将含IIDG 配电网看作不接入 IIDG 的配电网与接入的 IIDG 共 同构成的系统,则可以把含 IIDG 配电网短路电流计 算问题理解为:在不接入IIDG时的配电网短路电流 之上,叠加IIDG在发生故障时提供的短路电流,进 而得到含IIDG 配电网的短路电流。基于此,本文提 出将相同运行方式下发生相同故障但不接入IIDG 时测量点流过的短路电流L作为样本特征之一。由 于L是由不含IIDG的配电网通过故障分析物理模型 计算出的短路电流,是从机理层面提出的样本特征, 不仅提高了机器学习模型的可靠性,而且由于其中 隐含着运行方式、故障特征等部分配电网信息,可极 大地减少特征数量。但是,研究发现,若旋转式电源 在配电网发生故障时不向测量点提供短路电流,即 L恒定为0,则该特征不包含任何信息;反之,当某些 运行方式下旋转式电源在发生故障时向测量点提供 短路电流,即I_r不恒定为0,则该特征具有物理意义 和包含配电网信息,因此I_r这一特征仅适用于I_r不恒 定为0,即相应机理有效的情况。则样本组成见表1。

表1 样本组成

Table 1 Composition of data set

样本组成	符号	变量含义				
特征I	$I_{ m f}$	不接入IIDG时的短路电流				
	V_m θ_m	节点m的电压幅值、相角				
桂毎Ⅱ	Z_i	旋转式电源 i 的等值阻抗				
	P_{linen} Q_{linen}	线路n传输的有功、无功功率				
(檍态运行时)	$P_{{\scriptscriptstyle \mathrm{L}}k}$, $Q_{{\scriptscriptstyle \mathrm{L}}k}$	负荷节点k的有功、无功功率				
	$I_{\mathrm{DG}j}$	第 <i>j</i> 个IIDG注入电流				
	α_{j}	第j个IIDG投入配电网的情况				
共同特征Ⅲ	$S_{\mathrm{DG}i}$	第j个IIDG投入配电网的容量				
	$A_{\rm line_cut}$	被切断线路集				
标签	$I_{\rm f_DG}$	接入IIDG时的短路电流				

表1中,特征 I 仅使用 I_i 作为关键特征;特征 II 是除共同特征 III 以外的其他配电网特征;共同特征 III 是所有样本共同需要的特征。特征 III 中,IIDG 投 入配电网的情况 α_j 、投入 IIDG 容量 S_{DG} 是反映 IIDG 对短路电流影响的特征, $\alpha_j=0$ 表示第j个 IIDG 不投 入配电网, $\alpha_j=1$ 则表示第j个 IIDG 投入配电网;被切 断线路集 $A_{\text{line_eut}}$ 是表示配电网拓扑结构变化的特征。 因此实际应用的特征为组合 I + III 或者组合 II + III, 其中组合 I + III 是包含机理的特征组合,组合 II + III 是常规配电网特征组合。接入 IIDG 后测量点流过 的短路电流 I_{LDG} 是样本标签。特征与样本标签共同 组成含 IIDG 配电网短路电流计算样本。

2 机器学习方法

短路电流是一个连续的量,因此,本问题在机器 学习领域属于回归问题。目前常用于回归问题的方 法包括线性回归方法、支持向量回归方法、集成学习 方法等。线性回归方法在面对非线性问题时学习效 果较差,支持向量回归方法对于大规模样本的计算 速度较慢,因此集成学习方法是更好的选择。集成 学习方法^[12]是目前运用非常广泛、学习效果出色的 机器学习方法,它的主要思想是训练多个弱学习器, 最后将这些弱学习器进行组合,通过投票(分类)或 取平均值(回归)的方式得到最终结果,即成为一个 优秀的强学习器。因此相对于其他方法,集成学习 方法具有准确度高、泛化能力强等特点。它可分为 Bagging和Boosting两大类。

Bagging是一种在原始数据集上通过抽样重新 选出新数据集来分别训练多个弱分类器进行集成的 方法,各个弱分类器是相互独立的,可以分开训练。 Bagging包括随机森林^[13]、极限随机树^[14]等。随机森 林算法是一种基于决策树的集成方法,通过随机抽 取样本,形成一系列子数据集来建立决策树,最后将 这些决策树进行组合。极限随机树算法和随机森林 算法相似,主要区别有2个:①每棵决策树通过完全 随机地选择分裂阈值完成节点分裂;②每棵决策树 均使用全部训练样本而非抽取训练样本子集生成 树。这样做的好处在于:从偏差-方差权衡的角度 看,完全随机的分裂节点生成决策树能够降低算法 的方差,采用全部样本进行训练则可以降低偏差;而 从计算速度的角度看,尽管建树的复杂度和其他基 于树的集成方法相同,但极限随机树算法在分裂节 点时不需要经过反复计算来决定分裂阈值,过程更 加简单,计算速度得到了提升,因此更适合用于大规 模数据的训练和学习。

Boosting则使用全部样本进行训练,各弱分类器的训练方式类似于迭代,下一轮训练的弱分类器模型将根据上一轮弱分类器模型的结果进行相应的调整,各个弱分类器间是顺序递进关系,有较强的关联关系。Boosting最主要的弱分类器模型是决策树,算法框架包括XGBoost^[15]、LightGBM^[16]等。XGBoost是一种基于梯度提升的算法,因其优良的学习效果在大数据竞赛中取得了优异表现,其特点是在损失函数中加入正则项,兼顾了模型的学习能力和泛化能力,同时在树节点分裂时可并行计算各特征最优划分,提高计算效率。LightGBM是微软亚洲研究院于2016年提出的另一个Boosting框架,对XGBoost的一些缺点进行了改进,主要特点是在不降低准确度的基础上极大地提升了计算速度。

上述4种算法,将用于下文的算例分析,以说明 集成方法可用于含IIDG 配电网的短路电流计算。

3 样本生成方法

机器学习建模依赖于能够反映特征与标签之间 映射关系的数量充足且完备的样本集合。对于短路 电流问题,当系统运行方式发生变化时,短路电流将 随之变化。因此参考文献[17],考虑采用以下方法 生成不同的系统运行方式,获取样本集。为使得到 的样本更具有多样性,对于一个包含c个旋转式电 源、s个IIDG节点、l个负荷节点的配电网, 拟考虑 M (M≥1)种基础运行方式,其中每一种基础运行方式 均可按照需求设定旋转型电源、IIDG、负荷配置参 数。以某一种基础运行方式为例,在这个基础运行 方式下,各旋转型电源的等值阻抗为Z_{hasei}(i=1,2,…, c),各IIDG出力为S_{DGbasei}(j=1,2,…,s),各负荷节点 的有功需求、无功需求分别为 P_{Lbasek} 、 Q_{Lbasek} 、 $(k=1,2,\cdots,$ l)。设 τ_i (*i*=1,2,...,*c*)、 β_i (*j*=1,2,...,*s*)和 ρ_{Pk} 、 ρ_{Ok} (*k*= 1.2.…,l)分别为旋转型电源、IIDG和有功无功负荷 在[-0.2,0.2]范围内独立产生的随机量,则可以通过 式(1)--(3)产生不同的旋转型电源等值阻抗、IIDG 出力和负荷需求,获得系统的不同运行方式。

$$Z_i = Z_{\text{basei}}(1 + \tau_i) \quad i = 1, 2, \dots, c$$
 (1)

$$S_{\text{DG}_j} = S_{\text{DGbase}_j} (1 + \beta_j) \quad j = 1, 2, \cdots, s$$
(2)

$$\begin{cases} P_{Lk} = P_{Lhasek} (1 + \rho_{Pk}) \\ Q_{Lk} = Q_{Lhasek} (1 + \rho_{Qk}) \end{cases} \quad k = 1, 2, \cdots, l$$
(3)

同时,考虑 IIDG 的投入运行情况,对每个运行 方式随机产生一个矩阵 A=[α₁ α₂ … α_s],其中α_j(j= 1,2,…,s)的定义见表1。总负荷与总出力之间的不 平衡情况可由系统电源进行补偿。此外考虑网络的 拓扑结构变化,从N-1原则的角度出发,在每一种运 行方式下从线路集合中随机产生1条线路作为开断 的线路A_{line_cut}。线路集合中包含配电网中可以被切 除的线路,用线路编号来表示,其中编号0表示不切 除任何线路,即完整的网络结构。需要指出的是, A_{line_cut}的备选线路集合与故障发生位置有关,由于故 障发生处下游线路的变化不会影响故障发生处上游 点流过的短路电流,因此需要根据设置的故障场景 生成A_{line_cut}的备选线路集合。

在产生样本时,首先通过上述运行方式的随机 设置可以得到特征Ⅲ,然后模拟出各个运行方式,设 置故障位置、故障类型和过渡电阻后,通过建模仿真 得到不接入任何 IIDG 情况下流过测量点的特征 I, 即*I*_r,或者得到接入实际 IIDG 后的样本特征 II,最后 通过故障仿真得到流经测量点的短路电流*I*_{LDC},形成 样本标签。由此可以积累起样本集合。

4 短路电流计算流程和评价指标

以特征组合 I +Ⅲ为例,说明基于集成方法建模的短路电流计算流程,流程图如附录中的图 A1 所示。整个流程可分为离线训练阶段和在线应用阶段。在训练阶段,首先设置 M 种基础运行方式的配置参数以及每种基础运行方式所需的样本数,在各基础运行方式的基础上,随机生成各个样本对应的运行方式,包括系统等值阻抗、负荷功率需求、网络拓扑等,通过仿真得到 I_t和 I_{cD6},结合特征 Ⅲ形成原始样本集;然后经过样本数据预处理过程,主要是对 A_{line_cut}进行独热处理,得到适合机器学习模型训练的 样本集;最后利用交叉验证选择出机器学习合适的 超参数进行训练,得到最终的机器学习模型。在应 用阶段,由给定的运行方式形成样本特征,经过和训 练阶段相同的数据预处理过程,使用训练好的模型 进行预测,即可得到最终的短路电流计算结果。

为衡量短路电流预测的准确性,选择平均绝对 百分比误差MAPE(Mean Absolute Percentage Error) 作为衡量机器学习准确性的指标,其定义为:

$$\lambda_{\text{MAPE}} = \frac{100}{N} \sum_{t=1}^{N} \frac{\left| y_t - \hat{y}_t \right|}{\left| y_t \right|} \tag{4}$$

其中,N为参与准确性评价的样本数量;y,为第t个样

本的标签; \hat{y}_i 为第t个样本的预测值。MAPE用于衡量短路电流预测值和实际值的相对误差,MAPE的值越小,说明短路电流预测越准确。

5 算例分析

5.1 算例及故障样本生成

为验证本文方法的可行性,采用IEEE 34节点 改进测试网络作为算例,该网络的结构图如图1所 示。配电网的电压等级为24.9 kV,其中共包含31条 线路,系统电源接在节点800即配电网络首端,同时 将IIDG并入该算例系统。由于此网络包含不平衡 馈线,包括单相馈线和三相馈线,且三相馈线相间互 阻抗各不相同,各相负荷也不相同,因此可作为不平 衡系统进行分析。



图 1 IEEE 34节点配电网结构图 Fig.1 Structure of IEEE 34-bus distribution system

在 MATLAB / Simulink 中搭建上述的 IEEE 34 节点配电网系统。考虑M=2,即系统最大、最小2种 基础运行方式,系统等值阻抗分别为j0.5 Ω 、j1 Ω 。 负荷与 IIDG 容量在 2 种基础运行方式中相同。其 中,负荷按照文献[18]中进行设置;配电网中的 IIDG 包括节点 844 接入基础容量为 500 kW 的 IIDG₁ 和 400 kW 的 IIDG₄,节点 832 接入 200 kW 的 IIDG₂, 节点 852 接入 500 kW 的 IIDG₃, IIDG 均具有故障穿越 能力,在配电网发生故障或异常时,逆变器的最大输 出电流均为其额定电流的 2 倍。在基础运行方式的 基础上,在[-0.2,0.2]变化范围内,随机产生系统等 值阻抗、负荷、IIDG 容量的变化量,以及 IIDG 投入情 况矩阵 A 和被切断线路 $A_{ling cut}$,生成新的运行方式。

在网络中设置短路故障,生成故障样本。故障 样本描述为:故障 f₁和 f₂分别设置在节点 834 和 860,故障类型均为三相短路,所求短路电流为故障 线路的首端,测量点分别为K₁、K₂和K₃;对每个故障 场景分别设置3000种运行方式,即每种基础运行方 式各设置1500组;通过仿真,提取出每种运行方式 下的特征和标签,各得到3000组样本,构成原始样 本集。将原始样本集按8:2的比例划分为训练集以 及测试集,即训练集2400组,测试集600组。由于 IEEE 34节点系统为三相不对称系统,因此,以下算例 仅选取c相电流作为样本标签进行分析,a、b相类似。 5.2 **算例分析**

首先说明模型的超参数选择过程,并使用特征 组合 I+Ⅲ进行不同算法的效果对比;然后对比2种 特征组合,分析机理与数据驱动融合模型和单纯的 数据驱动模型的效果;最后对样本数量进行讨论。

(1)超参数选择。

在建立机器学习模型时,为保证模型取得最优效果,需要对算法的超参数进行选择。最优超参数的确定通过网格搜索实现,其过程为:确定各超参数可选值,将所有的参数组合情况进行遍历,用训练集5折交叉验证的MAPE平均值作为评价标准,找到结果最优的参数组合,作为最终模型的超参数取值。

(2)不同算法的效果对比。

分别使用随机森林算法、极限随机树算法、XG-Boost算法、LightGBM算法,基于相同的训练样本集 建模,用相同的测试样本集进行测试,实现含 IIDG 的配电网短路电流计算。各个算法对测量点 K_1 、 K_2 、 K_3 流过短路电流的预测值与实际值之间的误差 如表 2 所示,表中包含各算法在训练集和测试集上 的表现,并用下划线标出最小值,后同。从 K_1 对应的 测试集中随机抽取出 5 组样本,进行各算法结果对 比,结果如附录中的表 A1 所示,表中 λ_{APE} 为绝对百 分比误差。

表2 不同算法预测结果对比

Table 2 Prediction error comparison of different models

)	1.01	
测量片	質法	Λ _{MAP}	_E / %	
例里尽	开拓	训练集	测试集	
	极限随机树	0.058	<u>0.253</u>	
V	随机森林	0.143	0.338	
\mathbf{K}_{1}	XGBoost	0.014	0.310	
	LightGBM	0.075	0.458	
	极限随机树	<u>0.078</u>	0.793	
V	随机森林	0.515	1.298	
κ ₂	LightGBM 极限随机树 施机森林 XGBoost LightGBM	0.113	0.514	
	LightGBM	0.381	0.516	
	极限随机树	0.066	<u>0.644</u>	
V	随机森林	0.350	0.897	
Λ ₃	XGBoost	0.041	0.745	
	LightGBM	0.208	0.741	

由表2可见,对于K₁、K₃,XGBoost在训练集上的 误差更小,而极限随机树方法相对于其他算法,在测 试集上的表现更好;对于K₂,极限随机树方法训练误 差更小,XGBoost则在测试集上的表现更好。再结 合表A1可以发现,4种算法的计算误差都在可以接 受的范围内,相比较而言随机森林方法的表现不突 出。因此,上述4种机器学习方法用于IIDG高渗透 配电网短路电流计算是可行的,计算误差较小。

(3)不同特征组合的机器学习模型对比。

对前述3000种运行方式,采用不同特征组合建

立机器学习模型,机器学习模型均采用极限随机 树算法。其中,特征组合Ⅰ+Ⅲ对应机理与数据驱动 融合模型,特征Ⅱ+Ⅲ对应单纯的数据驱动模型。表 3对比了2种特征组合对应模型的总体预测误差。

表3 不同特征组合的总体预测误差对比

 Table 3
 Prediction error comparison between different feature combinations

测量占	桂江卯厶	$\lambda_{\text{MAPE}} / \%$			
侧里吊	何仙祖百	训练集	测试集		
K_1	<u>I + III</u>	<u>0.058</u>	0.253		
	II + III	0.203	1.426		
<i>K</i> ₂	I +Ⅲ	0.078	0.793		
	II + III	1.16×10^{-4}	0.449		
<i>K</i> ₃	<u>I + III</u>	<u>0.066</u>	0.644		
	II + III	0.175	1.779		

特征组合 I+Ⅲ包含 35 个左右的特征,而特征 组合 II+Ⅲ则包含 450 个左右的特征。因为A_{line_eut}备 选集合受到短路点的影响有所不同,因此不同故障 场景的特征数量会有一些差异。采用特征组合 I+ Ⅲ的模型仅需经过最优超参数选择,而采用特征组 合 II+Ⅲ的模型需先进行特征选择,即特征降维,再 进行最优超参数选择,才能使模型达到最优效果。 对此,本文采用基于树模型的特征选择方法进行特 征降维。表4是不同特征组合对应模型对测试集中 随机抽取的具体样本的预测结果对比。

表4 不同特征	且合的具体预测结果对比
---------	-------------

 Table 4
 Prediction result comparison between different feature combinations

测量点	样本	实际值 / A	预测值	直 / A	$\lambda_{\rm APE}$ / %	
			I +Ⅲ	∏ + ∭	I +Ⅲ	∐ + Ⅲ
	1	240.85	241.72	241.95	0.364	0.459
K_1	2	35.56	35.39	35.02	0.486	1.532
	3	<u>251.53</u>	<u>251.27</u>	241.51	<u>0.100</u>	3.982
<i>K</i> ₂	1	<u>20.40</u>	20.51	20.45	0.550	<u>0.229</u>
	2	42.97	42.61	42.84	0.839	0.314
	3	27.36	27.27	27.43	0.342	0.230
<i>K</i> ₃	1	<u>41.80</u>	<u>41.98</u>	41.13	<u>0.430</u>	1.604
	2	236.32	235.63	237.40	<u>0.293</u>	0.457
	3	233.23	233.12	236.82	0.048	1.542

从表4、5可见, K_1 、 K_3 使用特征组合 I+II时测 试集误差更小,而 K_2 使用特征组合 I+III时虽然误 差不是特别大,但使用特征组合 I+III时的误差更 小。究其原因,发生短路故障 f_1 时,系统电源没有向 K_2 提供短路电流,因此对于 K_2 而言, I_1 恒为0,特征 I 失效,而特征 II依然有效,所以对于 K_2 ,使用特征组 合 II+III误差更小。而对于 K_1 、 K_3 ,若主干线上的线 路没有被切断,系统电源会提供短路电流, I_r 不恒为 0,特征 I有效,相对于数量庞大而显冗余的特征 II, 特征 I 质量更高,机器学习算法相当于是在 I_r 基础 上修正各 IIDG 接入与 I_{LDG} 之间的关系,因此机器学 习能够取得更好的效果。由此可知,当机理未失效 时,机理与数据融合驱动模型相对于纯粹数据驱动 模型有更好的预测效果,而数据驱动模型可作为机 理失效时的补充。

(4)样本数量分析。

图 2 为不同样本数量下 2 种模型在 5 折交叉验 证下的效果对比,衡量的是模型的泛化能力。由图 2可见,对于K₁和K₃,特征组合 I+Ⅲ对样本数量依 赖较小,而特征组合Ⅱ+Ⅲ的预测准确度则较为依赖 样本数量,误差随着样本数量的增加有较大幅度的 下降:从整体上看,使用特征组合Ⅰ+Ⅲ的精度要高 于特征组合Ⅱ+Ⅲ。因此,与数据驱动模型相比,机 理与数据融合驱动模型对样本数量的要求低,具有 更好的适应性。而对于K,,样本数量为500时2种特 征组合的误差比较接近,随着样本数量的增大,使用 特征组合Ⅰ+Ⅲ的误差大于特征组合Ⅱ+Ⅲ的误差。 这说明在机理失效的场景下,当样本数量较少时,机 器学习方法从2种特征组合中都难以学习到比较完 整的信息,当样本数量增加后,特征组合Ⅱ+Ⅲ所包 含的信息大于特征组合Ⅰ+Ⅲ,此时使用特征组合 Ⅱ+Ⅲ的机器学习模型,会具有更好的泛化能力。



图 2 不同样本数量下 2 种特征组合的结果对比 Fig.2 Result comparison between two feature combinations under different numbers of data sets

5.3 与物理建模方法对比

为进一步说明本文方法的有效性,选取文献[5] 中的算法进行对比。采用文献[5]中的IEEE 33节 点算例,该网络结构图如图3所示。本文方法与文献[5]方法计算结果的对比如表5所示,其中,本文方法对每个节点处故障训练一个极限随机树模型, 样本形成方法如前文所述,样本数量均为2500组。 需要说明的是,由于仿真中使用的模型(如系统电源、线路、负荷、IIDG等)可能与文献[5]存在不同, 因此本文得到的仿真结果与文献[5]有一些差别。



图 3 IEEE 33 节点配电网结构图

Fig.3 Structure of IEEE 33-bus distribution system



Table 5 Result comparison between proposed method and Reference $\lceil 5 \rceil$

短路点	本文方法的结果			文献[5]方法的结果		
位置	计算	仿真	误差 / %	计算	仿真	误差 / %
节点5	0.7496	0.7486	0.13	0.6244	0.6225	0.31
节点12	0.1650	0.1647	0.19	0.1595	0.1593	0.13
节点18	0.0555	0.0554	0.19	0.0399	0.0414	3.62
节点31	0.1564	0.1565	0.06	0.1500	0.1490	0.67

注:文献[5]中节点31的结果小数点疑标错,表中已修改。

由表5可以看出,与文献[5]中使用的物理建模 方法相比,本文采用机理与数据融合驱动建模方法 得到的短路电流可取得相近或更小的误差,验证了 所提方法的正确性。此外,在计算速度方面,机器学 习模型主要在模型训练上消耗时间,而预测需要的 时间很少,且与接入的 IIDG 数量及控制策略无关; 而传统物理建模方法,大多需经过多次迭代达到精 度要求后才停止计算,因此计算时间会受 IIDG 数量 影响。本节中4个模型,平均训练时间为2.235 s,平 均预测时间为0.049 s,因此在模型训练完成后,完全 能满足一些对计算速度要求很高的应用场景。

6 结论

(1)提出了一种基于机理与数据融合驱动的含 IIDG配电网短路电流计算新方法。通过IEEE 34节 点算例分析,以及与物理建模方法算例的对比,验证 了该方法应用于含IIDG配电网的短路电流计算时, 具有更准确、快速的优点。因此,本文所提方法可用 于自适应保护^[19]、故障管理等以短路电流计算为基 础的实时应用中。

(2)对比了2种配电网特征组合方式。特征组合 I+Ⅲ属于机理与数据融合驱动,使用不接入 IIDG 时的配电网短路电流,相当于将故障分析机理

加入机器学习模型,因此对于有系统电源流过的测量点具有较小的计算误差,且对样本数量的依赖性小;特征组合Ⅱ+Ⅲ则使用完整的配电网特征,因特征冗余需要进行特征选择,但为不流过系统电源电流的测量点提供了更为准确的计算模型,不足之处是对样本数量依赖性大。

(3)对机器学习中的不同集成算法进行了对比, 在最优超参数下误差水平均可接受,不同模型的计 算效果主要受到样本数量和特征组合影响。

然而本文的机器学习模型是在故障信息、IIDG 控制策略等已知的情况下建立的,是否可以把故障 信息,即故障位置、故障类型、过渡电阻等以及 IIDG 控制策略也作为模型输入,从而建立更通用的含 IIDG 配电网短路电流计算模型还需要进一步探索。

附录见本刊网络版(http://www.epae.cn)。

参考文献:

- [1]陈晓龙,李永丽,赵曼勇,等. 计及配电网电流保护整定原则的 逆变型分布式电源准入容量计算[J]. 电力自动化设备,2016, 36(6):156-162.
 CHEN Xiaolong,LI Yongli,ZHAO Manyong, et al. IBDG penetration level calculation considering current protection setting principle for distribution network[J]. Electric Power Automa-
- [2] 王成山,孙晓倩.含分布式电源配电网短路计算的改进方法
 [J].电力系统自动化,2012,36(23):54-58.
 WANG Chengshan,SUN Xiaoqian. An improved short circuit calculation method for distribution network with distributed generations
 [J]. Automation of Electric Power Systems, 2012,
- 36(23):54-58.
 [3] 潘国清,曾德辉,王钢,等.含PQ控制逆变型分布式电源的配电网故障分析方法[J].中国电机工程学报,2014,34(4):555-561.
 PAN Guoqing,ZENG Dehui,WANG Gang, et al. Fault analysis on distribution network with inverter interfaced distributed generations based on PQ control strategy[J]. Proceedings of the CSEE,2014,34(4):555-561.
- [4] WANG Qianggang, ZHOU Niancheng, YE Ling. Fault analysis for distribution networks with current-controlled three-phase inverter-interfaced distributed generators[J]. IEEE Transactions on Power Delivery, 2015, 30(3):1532-1542.
- [5]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.
- [6] 谈林涛,李军良,任昺,等.基于RB-XGBoost算法的智能电网 调度控制系统健康度评价模型[J].电力自动化设备,2020,40 (2):189-195.
 TAN Lintao,LI Junliang, REN Bing, et al. Health evaluation model of smart grid dispatch and control system based on RB-XGBoost algorithm[J]. Electric Power Automation Equipment,2020,40(2):189-195.
- [7] 鞠平,周孝信,陈维江,等. "智能电网+"研究综述[J]. 电力自动化设备,2018,38(5):2-11.
 JU Ping, ZHOU Xiaoxin, CHEN Weijiang, et al. "Smart grid plus" research overview[J]. Electric Power Automation Equipment,2018,38(5):2-11.

- [8]张沛超,陈琪蕾,李仲青,等.具有增量学习能力的智能孤岛检测方法[J].电力自动化设备,2018,38(5):83-89.
 ZHANG Peichao, CHEN Qilei, LI Zhongqing, et al. Intelligent islanding detection method with incremental learning capability
 [J]. Electric Power Automation Equipment,2018,38(5):83-89.
- [9] CHEN Y Q, FINK O, SANSAVINI G. Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction [J]. IEEE Transactions on Industrial Electronics, 2018, 65(1):561-569.
- [10] 王琦,李峰,汤奕,等. 基于物理-数据融合模型的电网暂态频率特征在线预测方法[J]. 电力系统自动化,2018,42(19):1-9.
 WANG Qi,LI Feng,TANG Yi,et al. On-line prediction method of transient frequency characteristics for power grid based on physical-statistical model[J]. Automation of Electric Power Systems,2018,42(19):1-9.
- [11] 张晨宇,王慧芳,叶晓君. 基于XGBoost算法的电力系统暂态 稳定评估[J]. 电力自动化设备,2019,39(3):77-83,89.
 ZHANG Chenyu, WANG Huifang, YE Xiaojun. Transient stability assessment of power system based on XGBoost algorithm[J]. Electric Power Automation Equipment,2019,39(3): 77-83,89.
- [12] 周志华. 机器学习[M]. 北京:清华大学出版社,2016:171-191.
- [13] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45 (1):5-32.
- [14] GEURTS P, ERNST D, WEHENKEL L. Extremely randomized trees[J]. Machine Learning, 2006, 63(1): 3-42.
- [15] CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//KDD 16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.]: ACM, 2016:785-794.
- [16] KE G, MENG Q, FINLEY T, et al. LightGBM: a highly effi-

cient gradient boosting decision tree[C]//Advances in Neural Information Processing Systems. New York, USA: Curran Associates Inc., 2017: 3146-3154.

- [17] WANG Huifang, ZHANG Chenyu, LIN Dongyang, et al. An artificial intelligence based method for evaluating power grid node importance using network embedding and support vector regression[J]. Frontiers of Information Technology & Electronic Engineering, 2019, 20(6):816-828.
- [18] IEEE Power Energy Systems Distribution System Analysis Subcommittee. Distribution test feeder[EB / OL]. (2010-09-17) [2020-02-05]. http://www.ewh.ieee.org / soc / pes / dsacom / testfeeders / index.html.
- [19] SHEN Shaofei, LIN Da, WANG Huifang, et al. An adaptive protection scheme for distribution systems with DGs based on optimized Thevenin equivalent parameters estimation[J]. IEEE Transactions on Power Delivery, 2015, 32(1):411-419.

作者简介:



郑 翔(1994—),男,福建福州人,硕 士研究生,主要研究方向为电力系统数据挖 据(**E-mail**:281306045@qq.com);

王慧芳(1974—),女,浙江宁波人,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);

郑 翔

姜 宽(1993—),男,江苏徐州人,博 士研究生,主要研究方向为电力系统仿真与

继电保护(E-mail:jiangkuan@zju.edu.cn)。

(编辑 任思思)

Mechanism and data-driven combined short circuit current calculation method for distribution network with IIDG

ZHENG Xiang¹, WANG Huifang¹, JIANG Kuan¹, CAO Jianwei², LIU Ying², CHEN Yongwei²

(1. College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China;

2. State Grid Huzhou Power Supply Company, Huzhou 313000, China)

Abstract: The short circuit current in a distribution network with IIDGs (Inverter-Interfaced Distributed Generators) is mainly calculated by using physical modeling at present, which has drawbacks in terms of calculation speed, accuracy, and versatility in the case of high penetration of IIDG. A mechanism and data-driven combined short circuit current calculation method for distribution network with IIDGs and high penetration of IIDGs is proposed. To compare with simple data-driven calculation method, based on the feature analysis of distribution network with IIDGs, two kinds of feature combinations reflecting the short circuit current are introduced. In one of them, mechanism is considered and the short circuit current when IIDCs are not connected into the system is used as the key feature, while the other only uses features of the distribution network. The data sets are accumulated by running the simulation model on MATLAB / Simulink, and the machine learning models are trained with two feature combinations using the ensemble methods, which include random forest, extremely randomized trees, XGBoost and LightGBM. The feasibility and effectiveness of models based on ensemble methods are verified on the IEEE 34-bus system, and the calculation results of different ensemble methods and different feature combinations are compared. The results show that the ensemble learning methods are capable of correctly predicting the short circuit current, and the mechanism and data-driven combined method outperforms the simple data-driven method if the corresponding mechanism is still available. In addition, the comparison with the physical modeling method validates the accuracy and rapidity of the proposed method.

Key words: inverter-interfaced distributed generator; distribution network; short circuit current calculation; feature analysis; machine learning; ensemble methods 附录



图 A1 短路电流计算流程

Fig.A1 Flowchart of short circuit current calculation

表 A1 不同算法对 K1 测试集样本的预测结果对比

Table A1 Prediction result comparison among different algorithms on testing set of K_1

样本 实际值/A	极限随机树		随机森林		XGBoost		LightGBM		
	预测值/A	λ_{APE} /%	预测值/A	λ_{APE} /%	预测值/A	λ_{APE} /%	预测值/A	$\lambda_{APE}/\%$	
<u>1</u>	240.85	241.72	0.364	241.91	0.444	242.77	0.798	242.40	0.646
<u>2</u>	35.56	35.39	0.486	35.14	1.201	35.54	0.066	35.34	0.614
<u>3</u>	251.53	251.27	0.100	251.66	0.052	251.37	0.062	251.46	0.028
<u>4</u>	241.46	242.85	0.574	243.07	0.668	243.83	0.980	242.97	0.623
<u>5</u>	<u>24.16</u>	<u>24.15</u>	<u>0.063</u>	24.11	0.223	24.08	0.328	24.05	0.449
λ_{MAPE} /% 0.317		0.518		0.447		0.472			