

计及低频减载动作的最大暂态频率偏移快速估计

李常刚¹, 李华瑞¹, 刘玉田¹, 吴海伟², 徐春雷²

(1. 电网智能化调度与控制教育部重点实验室(山东大学), 山东省济南市 250061;

2. 国网江苏省电力有限公司, 江苏省南京市 210024)

摘要: 随着大容量远距离高压直流输电工程建设和大规模可再生能源的接入, 受端电网频率安全风险增大。针对大容量直流闭锁等可能触发低频减载的严重扰动, 文中提出基于机器学习的电力系统最大暂态频率偏移快速估计方法。将问题分解为低频减载响应判断和最大频率偏移估计两个子问题, 通过子模型交替求解估计最大暂态频率偏移; 基于支持向量回归方法构建最大频率偏移估计子模型, 以支持向量机为个体学习器构建基于 Bagging 集成学习的低频减载响应判断子模型; 以运行方式信息和扰动信息为输入, 采用 ReliefF 算法和主成分分析法对输入特征进行选择 and 提取, 降低模型复杂度。以某多直流馈入受端系统为例构建最大暂态频率偏移估计模型, 验证所提方法的准确性和适应性。

关键词: 电力系统; 频率偏移; 低频减载; 支持向量机; 集成学习; 特征降维

0 引言

大规模远距离输电是解决中国一次能源与负荷逆向分布矛盾的重要措施^[1], 大量特高压交直流工程的建设使得受端电网遭受严重功率扰动的风险增加。同时, 大量新能源发电的接入也使得系统惯性和调频能力下降, 受端电网发生严重频率偏移的风险增高^[2]。因此, 快速估计受端系统在大容量直流闭锁等严重功率扰动下的暂态频率偏移对于系统安全稳定运行是至关重要的。

目前常用的电力系统频率偏移分析方法可分为模型分析和机器学习两大类。模型分析法主要包括全时域仿真法和等值简化模型法。全时域仿真法能够计及电网详细模型, 但计算耗时, 难以满足在线应用要求; 等值简化模型由于忽略了电压动态或采用线性化假设等, 对大容量直流闭锁等严重功率扰动下的频率偏移分析准确性不足^[3-7]。

基于机器学习的系统频率偏移估计方法近些年得到广泛研究。文献[8-9]针对发电机脱网扰动, 以扰动量和扰动前机组出力、最大出力等为输入特征, 采用人工神经网络 (artificial neural network, ANN) 估计系统最低频率; 文献[10]以扰动量、扰动

前稳态信息和扰动后瞬间机组功率变化等为输入特征, 使用函数连接型神经网络 (functional link network, FLN) 估计系统最低频率; 文献[11]选择与文献[10]相似的输入特征, 基于 ν 支持向量回归 (ν -support vector regression, ν -SVR) 估计发电机脱网后的最低频率; 文献[12]基于支持向量回归 (support vector regression, SVR) 和 ANN, 对扰动后的稳态频率偏移量进行了预测; 文献[13]基于多层极限学习机, 对扰动后的极值频率、最大频率变化率和准稳态频率进行了预测。

上述机器学习模型拓展了暂态频率偏移估计的手段, 但都没有考虑低频减载 (under-frequency load shedding, UFLS) 动作对暂态频率偏移的影响。对于可能触发 UFLS 动作的扰动, 如大容量直流闭锁, 现有方法难以准确估计系统暂态频率偏移。

针对计及 UFLS 的系统最大暂态频率偏移估计问题, 本文将问题分解为 UFLS 响应判断和最大暂态频率偏移估计两个子问题, 分别基于集成学习和 SVR 构建机器学习子模型, 通过两个子模型交替求解, 快速估计系统最大暂态频率偏移。为降低模型复杂度, 提高模型适应性, 本文采用特征类别初筛、特征选择和特征提取技术实现模型输入特征降维, 并通过算例系统验证方法的准确性和适应性。

1 最大暂态频率偏移估计模型结构

1.1 问题分析

当受端系统发生直流闭锁等严重功率扰动时,

收稿日期: 2018-06-30; 修回日期: 2019-01-13。

上网日期: 2019-04-11。

国家重点研发计划资助项目 (2017YFB0902600); 山东大学青年学者未来计划 (2018WLJH31)。

频率快速下降。系统频率 f 变化由惯性方程决定:

$$T_j \frac{df}{dt} = P_m - P_e \quad (1)$$

式中: T_j 为系统总惯性时间常数; P_m 为与原动机-调速系统动态响应特性有关的系统机械功率; P_e 为系统电磁功率, 由三部分构成

$$P_e = P_0 + P_d - P_{\Sigma s} \quad (2)$$

式中: P_0 为扰动前初始负荷; P_d 为扰动量; $P_{\Sigma s}$ 为 UFLS 实际总切负荷量。

$P_{\Sigma s}$ 为系统 UFLS 各轮次所切负荷量 P_{si} 之和, 具体切负荷量与 UFLS 实际动作轮数 n_s 有关, 而 n_s 又取决于系统频率动态过程, 即

$$\begin{cases} P_{\Sigma s} = \sum_{i=1}^{n_s} P_{si} \\ n_s = g(f) \end{cases} \quad (3)$$

式(1)~式(3)共同描述了系统的频率动态过程。可知, 频率响应过程与 UFLS 动作互为因果, 是相互耦合的。

式(1)和式(2)的原动机-调速系统动态与惯性响应是连续过程, 而式(3)的 UFLS 响应则是离散事件。因此, 在直流闭锁等导致 UFLS 动作的扰动下, 对扰动后的系统最大暂态频率偏移进行估计, 是典型的离散-连续混合系统动态分析问题。

1.2 最大暂态频率偏移估计模型的分解建模

由于离散系统与连续系统的响应特性存在较大差异, 在离散-连续混合系统分析中, 一般将离散系统和连续系统分解后分别建模。

在估计系统暂态频率偏移时, 针对式(1)和式(2)描述的连续系统构建最大频率偏移估计子模型 F, 分析不考虑后续 UFLS 动作时系统的动态频率响应; 针对式(3)描述的离散系统构建 UFLS 响应子模型 U, 判断 UFLS 动作情况。F 和 U 与实际电力系统频率分析模型的关系如图 1 所示。

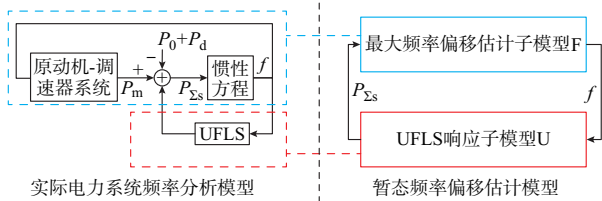


图 1 频率偏移估计模型与实际频率分析模型的关系
Fig. 1 Relationship between frequency deviation estimation model and actual frequency analysis model

当 UFLS 动作切除不同轮次负荷时, 除系统等效扰动量发生变化外, 负荷频率特性与电压特性也会发生变化, 进而影响系统频率响应。因此, 对于一个有 n 轮 UFLS 的电力系统, 本文分别建立判断第

i 轮 UFLS 是否动作的子模型 U_i 和分析第 i 轮 UFLS 动作而第 $i+1$ 轮不动作时系统频率偏移的子模型 F_i 。特别地, 当 $i=0$ 时, 子模型 F_0 分析系统所有 UFLS 轮次均不动作时的系统暂态频率偏移。

1.3 训练样本集生成

训练 F 和 U 子模型前, 需要首先生成相应的训练样本集。由于子模型 U_i 和 F_i 的输入输出及应用场景不同, 二者的训练样本集需要分别生成。

子模型 F_i 用于分析第 i 轮 UFLS 动作而第 $i+1$ 轮 UFLS 不动作时的系统最大暂态频率偏移, 需选择触发前 i 轮 UFLS 而不会触发第 $i+1$ 轮 UFLS 的样本, 即各个子模型 F 的训练样本集应满足式(4)和式(5)所示的关系。

$$X_{F_i} \cap X_{F_j} = \emptyset \quad 0 \leq i < j \leq n \quad (4)$$

$$X_{F_0} \cup \dots \cup X_{F_i} \cup \dots \cup X_{F_n} = X_F \quad 0 < i < n \quad (5)$$

式中: X_F 为模型 F 的总训练样本集; X_{F_i} 和 X_{F_j} 分别为子模型 F_i 和 F_j 的训练样本子集; \emptyset 表示空集; n 为 UFLS 的总轮数。

子模型 U_i 用于判断第 i 轮 UFLS 是否动作, 其训练样本应包含触发第 i 轮 UFLS 动作和未触发第 i 轮 UFLS 的样本。特别地, 触发第 i 轮 UFLS 动作的样本还应包含触发第 $i+1$ 轮 UFLS 动作和未触发第 $i+1$ 轮 UFLS 两种情形。因此, 子模型 U_{i+1} 的训练样本应包含在子模型 U_i 的训练样本中, 不同 U 子模型的训练样本集应满足式(6)。

$$X_U \supset X_{U_1} \supset X_{U_2} \supset X_{U_3} \supset X_{U_n} \quad 1 \leq i < j \leq n \quad (6)$$

式中: X_U 为模型 U 的训练样本集; X_{U_i} 为子模型 U_i 的训练样本子集。

为得到满足上述要求的各子模型训练样本集, 本文按照图 2 所示流程生成训练样本。图中, S_{si} 为第 i 个运行方式下的稳态潮流量; S_{di} 为第 i 个运行方式下的有功扰动量, 即故障直流闭锁前的有功功率; N 为总的运行方式数; $f_{\min, ij}$ 为第 i 个训练样本在设置前 j 轮次 UFLS 的情况下, 有功扰动后系统的最低暂态频率; R_{ij} 为第 i 个训练样本第 j 轮 UFLS 的动作情况。 $R_{ij} = 1$ 表示第 i 个训练样本第 j 轮次 UFLS 动作; $R_{ij} = -1$ 表示第 i 个训练样本第 j 轮 UFLS 不动作; N_{XF_j} 为 X_{F_j} 中的样本数, N_{XU_j} 为 X_{U_j} 中的样本数。

图 2 中, 筛选第 j 轮 UFLS 动作而第 $j+1$ 轮 UFLS 不动作的样本, 选取相应稳态潮流信息、扰动信息和暂态最低频率构成训练样本加入子模型 F_j 训练样本集中; 进一步选取相应稳态潮流信息、扰动

信息和第 j 轮 UFLS 动作情况作为训练样本,加入到子模型 U_j 训练样本集中。

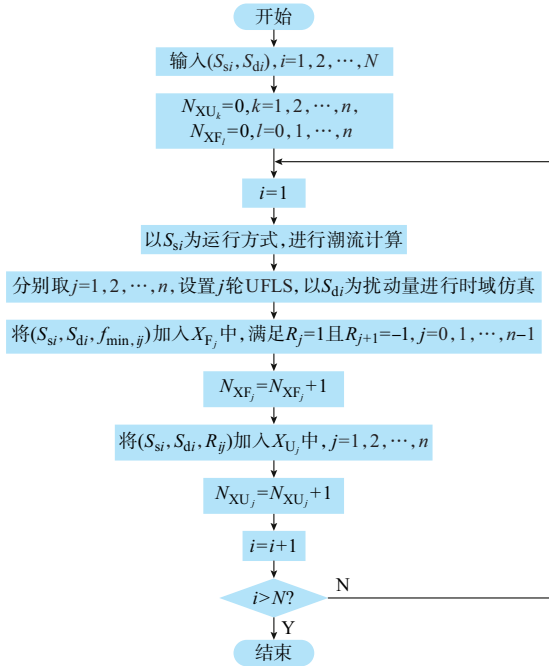


图2 训练样本集生成流程图

Fig. 2 Flow chart of generation for training sample set

1.4 最大暂态频率偏移交替估计流程

离散-连续混合系统一般采用离散系统模型和连续系统模型交替求解的方法进行响应特性分析。相应地,在构建子模型 F 和 U 后,本文也采用交替估计方法分析系统在直流闭锁等扰动后的暂态频率偏移,以反映频率偏移与 UFLS 动作的相互影响。子模型 F 与 U 的交替估计流程如图 3 所示。

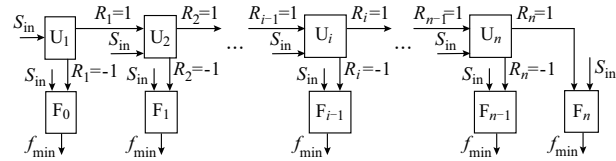


图3 系统最大暂态频率偏移交替估计流程图

Fig. 3 Flow chart of alternating estimation of system maximum transient frequency deviation

图 3 中, S_{in} 为模型的输入特征,将 S_{in} 输入模型 U_i ,判断第 i 轮 UFLS 是否动作。若动作,则输出 $R_i=1$,启动模型 U_{i+1} 并输入 S_{in} 判断第 $i+1$ 轮 UFLS 是否动作;若判断不动作,则输出 $R_i=-1$,启动模型 F_{i-1} 并输入 S_{in} ,预测最大暂态频率偏移。图中, f_{min} 为最低暂态频率偏移量。

对于判断为触发前 i 轮 UFLS 动作且第 $i+1$ 轮 UFLS 不动作的样本,图 3 中会依次触发模型 U_1, U_2, \dots, U_{i+1} ,由于 U_{i+1} 此时输出为 -1 ,因此

该样本最终 f_{min} 由子模型 F_i 进行估计。

2 最大暂态频率偏移估计子模型构建方法

2.1 基于 SVR 的子模型 F 构建

最大暂态频率偏移估计属于回归问题,常用的方法包括 ANN^[8] 和 SVR^[11] 等。与其他人工智能方法相比,SVR 方法成熟且模型结构简单,只有罚参数和核参数两项参数需要确定。因此,本文基于 SVR 方法构建子模型 F。

对于最大频率偏移估计子问题,设 m 个样本构成的集合为 $\{(x_i, f_{min,i}) | i=1, 2, \dots, m\}$, x_i 和 $f_{min,i}$ 分别为第 i 个样本的输入特征和最低暂态频率偏移量。

高维电力系统中最大暂态频率偏移估计问题具有较强非线性,使用线性判别函数构建子模型 F 会存在较大的误差。因此,需要引入非线性核函数 $K(x_i)$,将低维空间中非线性的最大暂态频率偏移估计问题映射到高维空间,然后在高维空间中构造最优超平面,提高估计精度。此时判别函数为:

$$F(x_i) = \omega^T \phi(x_i) + b \quad (7)$$

式中: $\phi(x_i)$ 为样本 x_i 的高维映射函数; ω 为各维度特征权重列向量; b 为偏置量。

构建 SVR 模型的回归函数,需求解参数 ω 和 b 。可以将对参数 ω 和 b 的求解转化成以分类间隔最大为目标的优化问题,并引入拉格朗日乘子把凸优化问题简化为最大化二次型,即可通过解相应的二次规划问题求解^[14]。

常用支持向量机 (support vector machine, SVM) 核函数有径向基核函数 (radial basis function, RBF)、Sigmoid 核函数和多项式核函数等。当样本数目较大时, RBF 和 Sigmoid 核函数预测结果较好,二者收敛性优于多项式核函数;而当输入向量维数较高时, RBF 和多项式核函数的分类和回归效果优于 Sigmoid 核函数。由于 RBF 在不同样本规模和输入向量维数下都具有较高的非线性映射精度,收敛域宽,故本文选择 RBF 为子模型 F 的核函数,其结构为:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|_2^2} \quad (8)$$

式中: γ 为核半径。

2.2 基于 Bagging 集成学习的子模型 U 构建

在所构建的连续-离散混合模型中, UFLS 轮数判断错误会导致最大暂态频率偏移量的估计值出现较大误差。因此,提高 UFLS 动作轮数判断的准确性是提高频率偏移估计准确性的关键。由于单一学习器对复杂函数的表征能力有限,泛化能力不足,本文采用集成算法将若干独立的个体学习器综合为一

个集成学习器,提高估计精度^[15]。

2.2.1 基于 SVM 的 Bagging 集成学习模型

Bagging 算法是并行式集成学习方法的代表^[16-17],分为同质个体学习器集成和异质个体学习器集成。本文采用同质个体学习器集成的方法,构造 m_L 个不同的个体学习器, m_L 为奇数,并通过式(9)得到集成学习模型的最终输出结果。

$$H_i(\mathbf{x}_j) = \text{sgn}\left(\sum_{k=1}^{m_L} h_{ik}(\mathbf{x}_j)\right) \quad (9)$$

式中: $H_i(\mathbf{x}_j)$ 为输入样本为 \mathbf{x}_j 时,子模型 U_i 的最终预测输出结果; $h_{ik}(\mathbf{x}_j)$ 为输入样本为 \mathbf{x}_j 时,各个个体学习器的预测输出结果,1 和 -1 分别表示第 i 轮 UFLS 动作和不动作; $\text{sgn}(\cdot)$ 为取符号运算,参数为正值时返回 1,参数为负值时返回 -1。

模型 U 本质上属于分类问题,可采用决策树(decision tree, DT)^[18]、SVM^[14]等构建个体学习器。本文使用 SVM 构造 UFLS 响应判断的个体学习器,并采用与基于 SVR 的最大暂态频率偏移估计子模型相似的方法训练 SVM 个体学习器。

2.2.2 个体学习器的训练

集成学习方法的精度与其各个个体学习器的多样性有关。在构造个体学习器时,应对其训练样本集合的输入特征集进行随机选取,以保证各个个体学习器在评估结果上的多样性。

对于子模型 U_i 的第 j 个个体学习器 h_{ij} ,使用 Bootstrap 重采样的方法随机抽样生成其训练样本集和输入特征集,抽样比例均为 0.5,可以得到个体学习器的训练样本集 $X_{U_{ij}}$ 和输入特征集 $A_{U_{ij}}$,有 $X_{U_{ij}} \subset X_{U_i}$, $A_{U_{ij}} \subset A_{U_i}$, $j=1, 2, \dots, m_L$,其中 $A_{U_{ij}}$ 为个体学习器 h_{ij} 的输入特征集, A_{U_i} 为子模型 U_i 的输入特征集。

在使用 Bootstrap 方法随机抽样生成各个 SVM 个体学习器训练样本集时,可能存在数据集不平衡问题。如在子模型 U_i 的训练子集中,第 i 轮 UFLS 动作的样本数可能远远大于第 i 轮 UFLS 不动作的样本数。分类学习器以经验风险最小作为模型的学习准则,即学习目标是 minimized 模型在训练样本上分类误差,训练样本集的数据不平衡将会使训练过程提前结束,导致各个个体学习器泛化性能较差,影响最终集成学习器的估计精度。

为提高个体学习器的估计精度,解决个体学习器训练样本集中可能存在的数据不平衡问题,本文定义不平衡分类评价指标 I_{uc} 作为分类器在不平衡数据集上分类性能的评价指标, I_{uc} 值越大则分类器分类效果越好,其计算公式为:

$$I_{uc} = \sqrt{I_{acc+} + I_{acc-}} \quad (10)$$

式中: I_{acc+} 为正类分类精度; I_{acc-} 为负类分类精度,计算公式分别如式(11)和式(12)所示。

$$I_{acc+} = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (11)$$

$$I_{acc-} = \frac{N_{TN}}{N_{TN} + N_{FP}} \quad (12)$$

式中: N_{TP} 为预测正确的正类样本数; N_{FN} 为错误判断为负类的正类样本数; N_{TN} 为预测正确的负类样本数; N_{FP} 为错误判断为正类的负类样本数。

二分类问题中,正类通常指少数类样本,负类通常指多数类样本。指标 I_{uc} 只有在正类分类精度和负类分类精度都较高时,才能取得较大值。

设定阈值 $I_{uc,th}$,对于构成子模型 U_i 的各个个体学习器的训练子集 $X_{U_{ij}}$,如有第 i 个个体学习器不平衡分类评价指标 $I_{uc,ij} < I_{uc,th}$,则表明该训练子集的样本存在明显不平衡,可进一步使用合成少数过采样技术(synthetic minority oversampling technique, SMOTE)^[19]训练子集 $X_{U_{ij}}$ 中不同类别的数据,具体流程如下。

1)对每个少数类样本 \mathbf{x}_i ,搜索 k 个少数类近邻样本。

2)从 k 个最近邻样本中随机选择 L 个样本,记为 $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iL}$ 。

3)由 \mathbf{x}_i 和 $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iL}$,通过式(13)生成新的少数类样本 $\mathbf{x}_{n-i1}, \mathbf{x}_{n-i2}, \dots, \mathbf{x}_{n-iL}$ 。

$$\mathbf{x}_{n-ik} = \mathbf{x}_i + (\mathbf{x}_{ik} - \mathbf{x}_i)r \quad (13)$$

式中: r 为 0~1 之间的随机数。

将新生成少数类样本 $\mathbf{x}_{n-i1}, \mathbf{x}_{n-i2}, \dots, \mathbf{x}_{n-iL}$ 加入训练样本子集 $X_{U_{ij}}$ 中,当 $I_{uc,ij} > I_{uc,th}$,即可采用 $X_{U_{ij}}$ 训练 h_{ij} 。

通过上述方法,可解决各个个体学习器训练样本子集数据不平衡的问题,提高各个子模型的泛化性能,进而提高最终的集成学习模型的预测精度。

3 模型输入特征降维

频率响应估计模型的输入特征一般可分为三类:反映运行方式的扰动前潮流信息 S_s 、反映直流功率的扰动信息 S_d 和反映扰动后响应的动态信息 S_D 。获取扰动后的动态信息需采用时域仿真,会影响模型估计速度。为满足评估快速性要求,本文选择 S_s 和 S_d 为输入特征,以避免对时域仿真的依赖。随着电网规模增加, S_s 的维数增高会导致“维度灾”,需对 S_s 进行特征降维,以降低模型复杂度。

3.1 输入特征类别初筛

对于模型 F 和子模型 U 输入特征类别进行初

筛,以快速筛除与最大暂态频率偏移估计和 UFLS 响应判断关联性低的特征类别,具体方法如下。

1)根据专家经验初选 9 类变量为候选特征集:扰动前各直流功率 S_{DC} 、扰动前各支路有功功率 S_{line} 、扰动前各节点有功负荷 S_{load} 、扰动前各机组有功出力 S_{gen} 、扰动前系统总负荷 $S_{\Sigma load}$ 、扰动前机组总有功出力 $S_{\Sigma gen}$ 、系统总惯性系数 $S_{\Sigma H}$ 、系统总一次调频系数 $S_{\Sigma KG}$ 和系统总旋备容量 $S_{\Sigma Pmax}$ 。

2)候选特征类筛选。不同的初始候选特征类的组合会影响模型的估计精度,因此,需要对输入特征类组合方案 S_F 性能进行评估,使得最大暂态频率偏移量的估计误差均方根 σf_{min} 最小,即

$$\min \sigma f_{min}(S_F) \quad (14)$$

上述候选特征类别为 9 类,本文采用枚举法遍历所有候选特征集组合,根据频率偏移估计误差选择最优特征类组合。

3.2 子模型 F 输入特征选择

电力系统是复杂的高维非线性系统。如果将初筛得到的原始输入特征全部作为暂态频率偏移估计模型的输入,冗余特征会增加模型训练和估计时间,并可能导致估计精度下降。因此,需要对初筛得到的原始输入特征进行降维。

在使用 Bagging 集成算法构建 UFLS 响应判断子模型 U 时,需要对原始特征集进行采样以构成多个个体学习器的特征集,因此应保证子模型 U 原始特征集的多样性。因此,本文仅在构成子模型 F 的输入特征集时,对原始特征集进行特征选择;而在构建子模型 U 时,不进行特征选择。

特征选择直接从全部原始特征中选择最相关的特征子集,以降低样本维度,属于物理降维。常用的特征选择方法有混合互信息(hybrid mutual information, HMI)法^[20]、蚁群优化(ant colony optimization, ACO)算法^[21]和 ReliefF 算法^[22]等。

ReliefF 算法根据特征在同类和非同类的 k 个最近邻样本的距离计算特征 u_j 的权重 w_j ,是基于特征赋权的特征选择方法,具有计算量较小、适应性强的特点。本文使用 ReliefF 算法对初筛得到的扰动前潮流特征进行赋权,进而选择权重较大的特征作为输入特征。

假定经上节特征类初筛后的系统特征维数为 p ,样本 x_i 可表示为:

$$\mathbf{x}_i = [u_{i1} \quad u_{i2} \quad \cdots \quad u_{ip}] \quad (15)$$

式中: u_{ij} 为第 i 个样本的第 j 个输入特征的取值。

采用 ReliefF 算法计算特征 u_j 权重的方法为:

$$w_j = \frac{1}{mk} \sum_{i=1}^m D_j^i \quad (16)$$

$$D_j^i = - \sum_{l=1}^k d_j^{i-l}(\mathbf{x}_i, \mathbf{x}_l^{b_i}) + \sum_{B \neq b_i} \frac{m_B}{m - m_{b_i}} \sum_{l=1}^k d_j^{i-l}(\mathbf{x}_i, \mathbf{x}_l^B) \quad (17)$$

$$d_j^{i-l}(\mathbf{x}_i, \mathbf{x}_l) = \frac{|u_{ij} - u_{lj}|}{\max u_j - \min u_j} \quad (18)$$

式中: D_j^i 为样本 x_i 对属性 u_j 的权重贡献; b_i 为样本 x_i 所属类别; B 为除 b_i 外的其他类别; m_{b_i} 和 m_B 分别为类别 b_i 和 B 的样本数; $d_j^{i-l}(\mathbf{x}_i, \mathbf{x}_l)$ 为样本 x_i 和 x_l 在 u_j 属性上的差异; $\max u_j$ 和 $\min u_j$ 分别为属性 u_j 在所有样本上的最大值和最小值; $x_l^{b_i}$ 和 x_l^B 分别为样本 x_l 在其同类 b_i 和非同类 B 中的第 l 个最近邻样本。

3.3 子模型 F 输入特征提取

特征提取是对原始特征进行线性或非线性变换,得到数量更少但更具有表达能力的新特征,属于数学降维。特征提取有主成分分析(principal component analysis, PCA)方法^[23]、独立成分分析(independent component analysis, ICA)方法^[24]、基于遗传算法和 k 近邻(k -nearest neighbor, KNN)算法^[25]的组合特征提取法等。对于经过特征初筛和特征选择后得到的输入特征集,本文采用 PCA 方法进一步进行特征提取,得到子模型 F 最终的输入特征集。与 3.2 节中分析相似,仅对子模型 F 的输入特征进行特征提取,不对子模型 U 的输入特征进行处理。

假定经上节特征选择后,样本特征维数降为 q 。利用 m 个训练样本,首先对特征 u_j 采用式(19)进行 Z-score 标准化,得到第 i 个样本经标准化后的第 j 个特征 v_{ij} ,并构造训练样本矩阵 $\mathbf{V} = (v_{ij})_{m \times q}$ 。其中元素 v_{ij} 为:

$$v_{ij} = \frac{u_{ij} - \mu_j}{\sigma_j} \quad (19)$$

式中: μ_j 为特征 u_j 的均值; σ_j 为特征 u_j 的标准差。

PCA 方法本质上是正交线性变换,变换矩阵由 \mathbf{V} 相关系数矩阵特征向量构成。 \mathbf{V} 相关系数矩阵 \mathbf{R} 为:

$$\mathbf{R} = \frac{1}{m} \mathbf{V}^T \mathbf{V} \quad (20)$$

计算 \mathbf{R} 的特征值 $\lambda_1, \lambda_2, \cdots, \lambda_q$ 和对应的特征列向量 $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_q$,且对任意 $i < j$ 都有 $\lambda_i \geq \lambda_j$ 。特征值 λ_i 反映了第 i 个主成分的贡献度。前 p 个主成分的累计贡献率可定义为:

$$c_p = \frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^q \lambda_i} \quad (21)$$

c_p 反映了前 p 个主成分承载的原 q 个特征所包含信息量的百分比。实际研究中可根据精度需要选择合理的贡献率阈值 c_{th} , 确定满足 $c_r > c_{th}$ 的最小 r , 即可将由 q 维原始特征构成的样本 $[u_{i1}, u_{i2}, \dots, u_{iq}]$ 降至由 p 维新特征构成的模型输入特征 $[h_{i1}, h_{i2}, \dots, h_{ip}]$:

$$\begin{bmatrix} h_{i1} \\ h_{i2} \\ \vdots \\ h_{ip} \end{bmatrix} = \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_p^T \end{bmatrix} \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iq} \end{bmatrix} \quad (22)$$

利用特征提取后的 p 维样本训练子模型 F , 即可得到最大暂态频率偏移估计模型。

4 算例分析

4.1 算例电网介绍

本文针对某省级受端电网构建最大暂态频率偏移估计模型并验证其有效性。该系统总负荷为 58 GW, 共有 3 条直流馈入, 总直流馈入为 20 GW, 其中 1 号和 2 号直流线路容量均为 8 GW, 3 号直流线路的容量为 4 GW。系统当前负荷水平下旋转备用约为机组最大出力的 15%。电网共配置 7 轮 UFLS, 动作频率依次为 49.25, 49.0, 48.75, 48.5, 48.25, 48.0, 47.75 Hz, 切负荷比例依次为 4%, 5%, 6%, 6%, 6%, 6% 和 6%, 各轮动作延时均为 0.2 s。

为构建模型样本集, 本文构建峰谷差率约为 40% 的全网负荷变化曲线, 分别构建服从正态分布的各节点负荷变化曲线和各直流功率曲线, 并按照全网网损最小计算最优潮流, 在不同负荷水平下, 进行全网随机停机, 保证系统旋转备用水平在 15% 左右, 得到 9 000 个不同负荷水平、不同直流运行状态和不同机组出力组合的运行方式样本。按照 1.3 节方法, 得到模型的样本集, 随机将其中 60% 样本作为训练样本, 其余作为测试样本。

本文以 1 号直流线路闭锁扰动为例, 验证本文所提方法的有效性。1 号直流线路闭锁, 不同运行方式和直流功率下第一轮 UFLS 均动作; 最严重情况下, 系统前 4 轮 UFLS 动作。因此, 需要分别建立模型 U_2, U_3 和 U_4 进行 UFLS 动作轮数的判断, 建立模型 F_1, F_2 和 F_3 进行最大暂态频率偏移的估计。

4.2 模型输入特征降维

通过 3.1 节方法分析模型输入特征类别组合, 选择由 6 类特征构成的输入特征类别组合中均方根误差最小的 7 个输入特征类别组合, 将其按模型预测精度由高到低进行排列, 并与 9 类输入特征全部输入(方案 8)的情况下的模型精度进行对比, 如表 1 所示。其中, 每一类别的输入特征包括属于该类别的所有特征变量, 如 S_{gen} 包括所有机组的有功出力, e_{RMSE} 为均方根误差。

表 1 不同输入特征组合下估计误差均方根
Table 1 Root mean square error with different input feature combinations

方案	S_{load}	S_{gen}	$S_{\Sigma load}$	S_{DC}	$S_{\Sigma gen}$	S_{line}	$S_{\Sigma H}$	$S_{\Sigma KG}$	$S_{\Sigma Pmax}$	e_{RMSE}/Hz		
										模型 F_1	模型 F_2	模型 F_3
1	✓	✓		✓			✓	✓	✓	0.025	0.019	0.023
2		✓	✓	✓			✓	✓	✓	0.068	0.079	0.069
3	✓			✓	✓		✓	✓	✓	0.068	0.081	0.071
4	✓			✓		✓	✓	✓	✓	0.160	0.200	0.150
5				✓	✓	✓	✓	✓	✓	0.170	0.220	0.160
6			✓	✓		✓	✓	✓	✓	0.190	0.210	0.150
7			✓	✓	✓		✓	✓	✓	0.250	0.340	0.210
8	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.022	0.018	0.021

综合比较表 1 中方案 1~7 和方案 8 可知, 方案 1 下模型 F_1, F_2 和 F_3 的均方根误差均较小, 精度与方案 8 相当。同时, 输入特征类别数为 7 或 8 的最优输入特征类别组合, 其精度与方案 1 相当, 但输入特征增加。因此, 本文选择方案 1 的 $S_{load}, S_{gen}, S_{DC}, S_{\Sigma Pmax}, S_{\Sigma KG}$ 和 $S_{\Sigma H}$ 共 110 维特征作为原始输入特征。

使用 3.2 节方法对 110 维输入特征进行特征选择, 并选择特征权重阈值为 0.5, 剔除特征权重低于

0.5 的 53 个特征后, 得到 57 维输入特征, 并采用 3.3 节方法进行特征提取。以 1 号直流线路闭锁下的模型 F_1 为例, 前 p 个主成分累计贡献率和模型估计误差与主成分个数 p 的关系如表 2 所示。其中, $e_{RMSE, p}$ 表示选择前 p 个主成分作为输入时, 模型估计的暂态最大频率偏移量的误差均方根。

由表 2 可知, 通过 PCA 方法将模型输入特征降至 8 维时, 前 8 个主成分累积贡献率为 95%, 均方根误差为 0.027 Hz。与未进行特征降维时的均方

根误差相比,特征维数为8时模型仍具有较高精度。若继续降低输入特征维数,模型估计误差会迅速增加。经验证,模型 F_2 和 F_3 同样在输入特征维数为8时取得较小误差;继续减少输入特征维数会导致模型估计误差急剧增加。因此,本文将模型 F_1 、 F_2 和 F_3 输入特征均通过特征提取降至8维。

表2 主成分个数与累计贡献率、误差均方根关系
Table 2 Relationship between numbers of principal components, accumulative contribution rate and root mean square error

p	$c_p/\%$	$e_{\text{RMSE},p}/\text{Hz}$
4	28	0.470
5	39	0.360
6	64	0.190
7	93	0.029
8	95	0.027
57	100	0.025

4.3 模型训练结果分析

对于1号直流线路闭锁故障,在装有Intel® E5-2640 v3 @2.60 GHz CPU的服务器上,对模型 U_2 、 U_3 和 U_4 进行训练,模型训练误差及训练耗时 t_{train} 如表3所示。表中, N_F 为误判样本数,指实际不触发相应轮次UFLS而本文模型误判为触发相应轮次UFLS动作的样本个数; N_L 为漏判样本数,指实际触发相应轮次UFLS动作而本文模型误判为不触发相应轮次UFLS动作的样本个数。

表3 1号直流线路闭锁故障模型U训练结果
Table 3 Training results of model U when DC line 1 is blocked

模型	子学习器个数	训练样本数	N_F/N_L	$t_{\text{train}}/\text{s}$
U_2	51	5 400	3/2	334
U_3	51	2 425	5/4	279
U_4	51	687	1/1	92

由表3可知,1号直流线路闭锁故障的UFLS动作轮数判断子模型总训练时间约为12 min。如对模型 U_2 、 U_3 和 U_4 并行训练,可以进一步降低训练时间。训练样本上模型误差较小,最大误判率和漏判率分别为0.2%和0.16%。分别对模型 F_1 、 F_2 和 F_3 进行训练,模型训练误差如表4所示。其中, $\Delta f_{\text{min,max}}$ 为最大频率偏移训练误差。

表4 1号直流线路闭锁故障模型F训练结果
Table 4 Training results of model F when DC line 1 is blocked

模型	训练样本数	$\Delta f_{\text{min,max}}/\text{Hz}$	$e_{\text{RMSE}}/\text{Hz}$	$t_{\text{train}}/\text{s}$
F_1	2 312	0.021	0.008 3	260
F_2	2 325	0.023	0.007 5	262
F_3	763	0.018	0.005 2	90

由表4可知,1号直流线路闭锁故障最大暂态频率偏移子模型F总训练时间约为10 min,并行训练可进一步降低训练耗时。最大频率偏移训练误差最大为0.023 Hz,均方根误差不超过0.01 Hz。

4.4 模型泛化能力分析

使用测试样本测试模型 U_2 、 U_3 和 U_4 的泛化能力,误差如表5所示,其中 t_{test} 为测试所用时间。

表5 1号直流线路闭锁故障模型U测试误差
Table 5 Testing error of model U when DC line 1 is blocked

模型	子学习器个数	训练样本数	N_F/N_L	t_{test}/s
U_2	51	3 600	7/12	5.76
U_3	51	1 950	12/5	3.12
U_4	51	531	3/0	0.95

由表5可知,模型 U_2 、 U_3 和 U_4 的测试误差较训练误差略有增加,但模型 U_3 漏判率最大为0.62%,模型 U_2 误判率最大为0.33%,表明所训练模型具有较高的估计精度,泛化能力较强。

使用测试样本测试模型 F_1 、 F_2 和 F_3 的泛化能力,可得模型测试误差如表6所示。

表6 1号直流线路闭锁故障模型F测试结果
Table 6 Testing error of model F when DC line 1 is blocked

模型	训练样本数	$\Delta f_{\text{min,max}}/\text{Hz}$	$e_{\text{RMSE}}/\text{Hz}$	t_{test}/s
F_1	1 483	0.042	0.025	2.52
F_2	1 234	0.045	0.019	2.10
F_3	883	0.033	0.023	1.50

由表6可知,模型 F_1 、 F_2 和 F_3 的测试误差较训练误差略有增加,最大频率偏移训练误差为0.045 Hz,均方根误差不超过0.03 Hz,表明所训练模型具有较高的估计精度和泛化能力。综合上述分析, $\Delta f_{\text{min,max}}$ 均出现在子模型U发生误判或者漏判的样本中。本文通过Bagging集成学习方法,有效提高了UFLS判断精度及最大暂态频率偏移准确度。

4.5 与其他频率偏移估计方法的比较

为进一步分析本文所提频率偏移估计模型的适应性,本文以1号直流线路闭锁为例,将本文所建模型与其他文献中的方法进行比较,相关结果如表7所示。其中,方法1为ANN模型^[8-9],方法2为FLN模型^[10],方法3为 ν -SVR模型^[11]。

由表7可知,方法1~3由于未考虑UFLS动作影响,对于UFLS动作样本,估计误差较大。而本文方法及频率偏移与UFLS的耦合,能够准确评估UFLS动作样本,具有更好的适应性。

表 7 不同最大暂态频率偏移估计方法结果对比
Table 7 Results comparison of different estimation methods for maximum transient frequency deviation

预测方法	触发一轮 UFLS		触发两轮 UFLS		触发三轮 UFLS	
	$\Delta f_{\min, \max}/\text{Hz}$	$e_{\text{RMSE}}/\text{Hz}$	$\Delta f_{\min, \max}/\text{Hz}$	$e_{\text{RMSE}}/\text{Hz}$	$\Delta f_{\min, \max}/\text{Hz}$	$e_{\text{RMSE}}/\text{Hz}$
本文	0.042	0.025	0.045	0.019	0.033	0.023
1	0.160	0.110	0.350	0.150	0.460	0.190
2	0.120	0.090	0.240	0.130	0.320	0.240
3	0.170	0.140	0.270	0.210	0.340	0.250

5 结语

针对直流闭锁等严重功率扰动,本文提出基于机器学习的最大暂态频率偏移估计方法,构建了基于 Bagging 集成学习的 UFLS 响应判断和基于 SVR 的最大频率偏移估计子模型,由两个子模型交替求解实现计及 UFLS 离散动作的暂态频率偏移估计。算例分析表明,本文模型具有较好适应性,频率偏移估计精度高,UFLS 动作轮数错判率低,可用于大规模交直流混联受端系统频率安全风险快速评估和事故筛选。本文方法可进一步用于研究快速频率评估指导下的 UFLS 优化和频率安全预防控制等问题。

参考文献

- [1] 李明节. 大规模特高压交直流混联电网特性分析与运行控制[J]. 电网技术, 2016, 40(4): 985-991.
LI Mingjie. Characteristic analysis and operational control of large-scale hybrid UHV AC/DC power grids[J]. Power System Technology, 2016, 40(4): 985-991.
- [2] 李卫东, 晋萃萃, 温可瑞, 等. 大功率缺失下主动频率响应控制初探[J]. 电力系统自动化, 2018, 42(8): 22-30. DOI: 10.7500/AEPS20170930010.
LI Weidong, JIN Cuicui, WEN Kerui, et al. Active frequency response control under high-power loss [J]. Automation of Electric Power Systems, 2018, 42(8): 22-30. DOI: 10.7500/AEPS20170930010.
- [3] CHAN M L, DUNLOP R D, SCHWEPPE F. Dynamic equivalents for average system frequency behavior following major disturbances[J]. IEEE Transactions on Power Apparatus and Systems, 1972, PAS-91(4): 1637-1642.
- [4] ANDERSON P M, MIRHEYDAR M. A low-order system frequency response model[J]. IEEE Transactions on Power Systems, 1990, 5(3): 720-729.
- [5] 李常刚, 刘玉田, 张恒旭, 等. 基于直流潮流的电力系统频率响应分析方法[J]. 中国电机工程学报, 2009, 29(34): 36-41.
LI Changgang, LIU Yutian, ZHANG Hengxu, et al. Power system frequency response analysis based on the direct current load flow[J]. Proceedings of the CSEE, 2009, 29(34): 36-41.
- [6] EGIDO I, FERNANDEZ-BERNALI F, CENTENO P, et al. Maximum frequency deviation calculation in small isolated power systems [J]. IEEE Transactions on Power Systems, 2009, 24(4): 1731-1738.
- [7] MEDINA D R, RAPPOLD E, SANCHEZ O, et al. Fast

assessment of frequency response of cold load pickup in power system restoration[J]. IEEE Transactions on Power Systems, 2016, 31(4): 3249-3256.

- [8] KOTTICK D, OR O. Neural-networks for predicting the operation of an under-frequency load shedding system[J]. IEEE Transactions on Power Systems, 1996, 11(3): 1350-1358.
- [9] WU Yuankang. Frequency stability for an island power system: developing an intelligent preventive-corrective control mechanism for offshore location[J]. IEEE Industry Applications Magazine, 2017, 23(2): 74-87.
- [10] DJUKANOVIC M B, POPOVIC D P, SOBAJIC D J. Prediction of power system frequency response after generator outages using neural nets[J]. IEE Proceedings C-Generation, Transmission and Distribution, 1993, 140(5): 389-398.
- [11] 薄其滨, 王晓茹, 刘克天. 基于 v -SVR 的电力系统扰动后最低率预测[J]. 电力自动化设备, 2015, 35(7): 83-88.
BO Qibin, WANG Xiaoru, LIU Ketian. Minimum frequency prediction based on v -SVR for post-disturbance power system [J]. Electric Power Automation Equipment, 2015, 35(7): 83-88.
- [12] MOHAMMAD A, TURAJ A. Adaptive scheme for local prediction of post-contingency power system frequency [J]. Electric Power System Research, 2014, 107: 240-249.
- [13] 文云峰, 赵荣臻, 肖友强, 等. 基于多层极限学习机的电力系统频率安全评估方法[J]. 电力系统自动化, 2019, 43(1): 133-140. DOI: 10.7500/AEPS20180629012.
WEN Yunfeng, ZHAO Rongzhen, XIAO Youqiang, et al. Frequency safety assessment of power system based on multi-layer extreme learning machine [J]. Automation of Electric Power Systems, 2019, 43(1): 133-140. DOI: 10.7500/AEPS20180629012.
- [14] 畅广辉, 刘涤尘, 熊浩. 基于多分辨率 SVM 回归估计的短期负荷预测[J]. 电力系统自动化, 2007, 31(9): 37-41.
CHANG Guanghui, LIU Dichen, XIONG Hao. Short term load forecasting based on multi-resolution SVM regression[J]. Automation of Electric Power Systems, 2007, 31(9): 37-41.
- [15] XU Yan, ZHANG Rui, ZHAO Junhua. Assessing short-term voltage stability of electric power systems by a hierarchical intelligent system[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(8): 1686-1696.
- [16] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
ZHOU Zhihua. Machine learning [M]. Beijing: Tsinghua University Press, 2016.
- [17] WU Wenbin, PENG Mugen. A data mining approach combining k -means clustering with Bagging neural network for short-term wind power forecasting[J]. IEEE Internet of Things Journal, 2017, 4(4): 979-986.

- [18] 骆柏锋,穆云飞,贾宏杰,等.基于负荷特征库的大用户供电接入决策方法[J].电力系统自动化,2018,42(6):66-72. DOI: 10.7500/AEPS20170805001.
LUO Bofeng, MU Yunfei, JIA Hongjie, et al. Decision method of power supply access for large consumers based on load feature library [J]. Automation of Electric Power Systems, 2018, 42(6): 66-72. DOI: 10.7500/AEPS20170805001.
- [19] 黄建明,李晓明,瞿合祚,等.考虑小波奇异信息与不平衡数据集的输电线路故障识别方法[J].中国电机工程学报,2017,37(11):3099-3107.
HUANG Jianming, LI Xiaoming, QU Hezuo, et al. Method for fault type identification of transmission line considering wavelet singular information and unbalanced data set [J]. Proceedings of the CSEE, 2017, 37(11): 3099-3107.
- [20] 王皓,孙宏斌,张伯明.基于混合互信息的特征选择方法及其在静态电压稳定评估中的应用[J].中国电机工程学报,2006,26(7):77-81.
WANG Hao, SUN Hongbin, ZHANG Boming. Hybrid mutual information based feature selection method as applied to static voltage stability assessment in power systems [J]. Proceedings of the CSEE, 2006, 26(7): 77-81.
- [21] 王海燕,刘天琪,李兴源.基于蚁群优化算法的电力系统暂态稳定评估特征选择[J].电力系统保护与控制,2008,36(21):1-6.
WANG Haiyan, LIU Tianqi, LI Xingyuan. Feature selection based on ant colony optimization for power system transient stability assessment [J]. Power System Protection and Control, 2008, 36(21): 1-6.
- [22] 陈晓琳,姬波,叶阳东.一种基于ReliefF特征加权的R-NIC算法[J].计算机工程,2015,41(4):161-165.
CHEN Xiaolin, JI Bo, YE Yangdong. A R-NIC algorithm based on ReliefF feature weighting[J]. Computer Engineering, 2015, 41(4): 161-165.
- [23] 牟婷婷,陆微,王兰君,等.基于主成分分析的用电模式稳定性分析[J].电力系统自动化,2017,41(19):59-65. DOI:10.7500/AEPS20161222001.
MOU Tingting, LU Wei, WANG Lanjun, et al. Stability analysis of consumption mode based on principal component analysis [J]. Automation of Electric Power Systems, 2017, 41(19): 59-65. DOI: 10.7500/AEPS20161222001.
- [24] 司凤琪,周建新,仇晓智,等.独立成分分析方法在电站热力过程数据检验中的应用[J].中国电机工程学报,2008,28(26):77-81.
SI Fengqi, ZHOU Jianxin, QIU Xiaozhi, et al. Application of independent component analysis on the data validation of thermodynamic system in power plant [J]. Proceedings of the CSEE, 2008, 28(26): 77-81.
- [25] 管霖,王同文,唐宗顺.电网安全监测的智能化关键特征识别及稳定分区算法[J].电力系统自动化,2006,30(21):22-27.
GUAN Lin, WANG Tongwen, TANG Zongshun. Intelligent algorithm for kernel feature identification and stability-based system division in power grid security monitoring [J]. Automation of Electric Power Systems, 2006, 30(21): 22-27.

李常刚(1984—),男,通信作者,博士,副研究员,主要研究方向:电力系统运行与控制。E-mail: lichgang@sdu.edu.cn

李华瑞(1994—),男,博士研究生,主要研究方向:电力系统运行与控制。E-mail: lihuarui_sdu@163.com

刘玉田(1964—),男,博士,教授,博士生导师,主要研究方向:电力系统运行与控制。E-mail: liuyt@sdu.edu.cn

(编辑 鲁尔姣)

Fast Estimation of Maximum Transient Frequency Deviation Considering Under-frequency Load Shedding

LI Changgang¹, LI Huarui¹, LIU Yutian¹, WU Haiwei², XU Chunlei²

(1. Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education

(Shandong University), Jinan 250061, China;

2. State Grid Jiangsu Electric Power Co. Ltd., Nanjing 210024, China)

Abstract: With integration of high voltage direct current (HVDC) transmission with large capacity and long distance and large scale of renewable generation, the risk of frequency security at the receiving-end of power systems is rising. Aiming at severe disturbances such as large-scale HVDC blocking which may trigger under-frequency load shedding (UFLS), a method for fast estimation of maximum transient frequency deviation is proposed based on machine learning. The problem is decomposed into two sub-problems i. e. UFLS response judgment and maximum frequency deviation estimation, respectively. The maximum transient frequency deviation is estimated by solving the sub-models alternately. Support vector regression method is used to establish the sub-model of maximum frequency deviation estimation, the Bagging ensemble learning method based on support vector machine is used to establish the sub-model of UFLS response judgement. Operation condition and disturbance information are regarded as inputs. ReliefF method and principal component analysis are introduced to select and extract input features to reduce the model complexity. A receiving-end power system with multiple HVDC links is taken as an example to build a maximum transient frequency deviation model and verify the accuracy and adaptability of the proposed method.

This work is supported by National Key R&D Program of China (No. 2017YFB0902600) and Young Scholars Program of Shandong University (No. 2018WLJH31).

Key words: power system; frequency deviation; under-frequency load shedding (UFLS); support vector machine; ensemble learning; feature dimension reduction