

基于数据挖掘的区域暂态电压稳定评估

朱利鹏¹, 陆超², 孙元章¹, 黄河³, 苏寅生³, 李智欢³

(1. 武汉大学 电气工程学院, 湖北省 武汉市 430072;

2. 电力系统及大型发电设备安全控制和仿真国家重点实验室(清华大学), 北京市 海淀区 100084;

3. 中国南方电网电力调度控制中心, 广东省 广州市 510623)

Data Mining Based Regional Transient Voltage Stability Assessment

ZHU Lipeng¹, LU Chao², SUN Yuanzhang¹, HUANG He³, SU Yinsheng³, LI Zhihuan³

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

2. State Key Lab of Security Control and Simulation of Power Systems and Large Scale Generation Equipments (Tsinghua University), Haidian District, Beijing 100084, China;

3. Power Dispatching & Communication Center of China Southern Grid, Guangzhou 510623, Guangdong Province, China)

ABSTRACT: In allusion to the imperfection of the theory related to regional transient voltage stability assessment and insufficient reliability of engineering criteria, a data mining based method to assess regional transient voltage stability is proposed and a two-layer assessment framework, in which the interaction between single bus load stability and multi bus voltages is synthetically considered, is constructed. Utilizing the measure index of nodal stability and the identification based voltage/reactive power sensitivity matrix the original features of the power network is extracted. To cope with the difficult problem that there is not yet reliable standard to delimitate the regional transient voltage instability, a constraint based semi-supervised learning method is used to reliably classify and label the data sets. Based on decision tree algorithm a step-by-step updated classification model is established to generate the criterion of regional transient voltage stability, and by use of this classification model the inherent law related to the voltage partition and representative buses can be mined. The validity of the proposed assessment scheme as well as the adaptability and the accuracy of the classification model are verified by simulation results of EPRI 36-bus system.

KEY WORDS: regional transient voltage stability assessment; data mining; sensitivity identification; semi-supervised learning; decision tree

摘要: 针对区域暂态电压稳定评估相关理论还不完善、工程判据可靠性不足等问题,提出了基于数据挖掘的区域暂态电压稳定评估方法,构建了综合考虑单点负荷稳定与多点电压相互影响的2层评价框架。利用节点稳定度量指标及基于辨

识的电压无功灵敏度矩阵提取网络原始特征。面对区域暂态电压失稳尚无可靠界定标准的难题,采用基于约束的半监督学习方式对数据集进行可靠分类。基于决策树算法建立逐步更新的分类模型,生成区域暂态电压稳定判据,通过模型挖掘出有关电压分区、代表节点的内在规律。EPRI 36节点系统上的仿真结果证明了评估方案的有效性,以及分类评估模型的适应性和准确性。

关键词: 区域暂态电压稳定评估; 数据挖掘; 灵敏度辨识; 半监督学习; 决策树

DOI: 10.13335/j.1000-3673.pst.2015.04.023

0 引言

近年来世界范围内发生多起电压崩溃事故,使得电力系统电压稳定性问题得到越来越多的关注^[1-4]。短期大扰动所引起的暂态电压失稳或崩溃尤为严重^[5-8],在线监测时若能准确判断出区域性的暂态电压失稳并及时采取预防控制措施,将有利于遏制暂态电压失稳事故中部分节点所引起的大范围连锁性电压失稳或崩溃,提高区域整体稳定性。实际运行中常采用基于固定电压阈值的工程判据对暂态电压失稳进行判断,这类判据简便易行,但阈值的设定缺乏足够的理论支撑,其可靠性和选择性都难以保证^[9-10]。同时,区域暂态电压稳定评估的相关理论目前也还不甚完善,这给区域暂态电压稳定性的评估和判断带来了不少困难。

随着 WAMS/PMU 的迅速发展和推广,目前有不少学者研究利用数据挖掘和机器学习的方法进行电力系统在线监测和安全稳定评估。文献[11]利用决策树算法实现了系统主动解列的辅助决策。文献[12]提出结合 PMU 量测和决策树算法来实现系

基金项目: 国家重点基础研究发展计划资助项目(2012CB215206); 国家自然科学基金资助项目(51037002, 51107061)。

Project Supported by the National Basic Research Program of China (2012CB215206) and National Natural Science Foundation of China (51037002, 51107061).

统动态安全评估的方案。文献[13]基于概率神经网络和径向基函数的组合神经网络进行暂态稳定评估及裕度预测。文献[14]利用支持向量机增量学习方法来进行暂态稳定的在线评估。这类研究大多是在所研究问题的理论基础相对完善的前提下开展的,建立的模型主要用于分类评估,并不涉及对未知知识的探索发现。事实上,数据挖掘又称为知识发现(knowledge discovery in data, KDD),除能进行有效分类及模式识别外,还可挖掘出潜在有用的新知识。若已知部分先验知识,半监督学习方式能高效、可靠地引导探索式知识发现过程。

综上所述,本文提出基于数据挖掘的区域暂态电压稳定评估方案。首先构建从单点到区域的2层评价框架,提取节点稳定度量指标及基于辨识的电压无功灵敏度矩阵作为数据挖掘的原始特征变量,然后通过基于约束的半监督聚类学习对数据集进行划分,以此为基础,利用决策树算法建立逐步更新的分类模型,生成区域暂态电压稳定判据。最后通过EPRI36节点系统对评估方案进行验证,结合具体的分类模型对机理性规律进行归纳。

1 基本分析框架

感应电动机等动态负荷的特性是造成负荷节点暂态电压失稳的主要原因^[9-10,15-16],若诸多负荷节点的失稳最终导致整体性的电压跌落或崩溃,这时将引起区域暂态电压失稳。结合区域稳定性与负荷节点稳定性之间的密切联系,将区域暂态电压稳定评估分为2个层次,基本分析框架如下:

1) 从局部节点出发,根据广域测量信息,利用基于转矩平衡的判据^[9-10]得到区域内各等值负荷节点的稳定性。通过量化指标,把节点稳定性由传统的稳定、不稳定2类扩展为超稳定、稳定、弱稳定、临界稳定、失稳、严重失稳等子类别,为节点稳定程度提供更详细的描述。

2) 从区域整体出发,通过实测数据辨识得到反映区域网络内节点间相互影响的灵敏度矩阵,结合第1层的各节点稳定程度,提取出反映区域网络暂态特性的特征变量,根据数据挖掘所建立的分类评估模型对区域暂态电压稳定做出快速评估和判断。

2 原始特征变量提取

根据前述基本分析框架,本文从暂态过程中能突出反映区域网络各节点稳定程度及节点间相互影响关系的变量出发:利用综合负荷的动态转矩平衡关系,建立节点稳定度量指标;结合区域网络内节点间电压、无功的变化特性,实测辨识得到电压

无功灵敏度矩阵。此二者的集合即作为数据挖掘的原始特征集。

2.1 节点稳定度量指标

节点负荷模型采用感应电动机与ZIP静态负荷并联的综合负荷模型,其中感应电动机部分采用三阶机电暂态模型。滑差的变化间接反映电动机电磁转矩与机械转矩的动态平衡关系及电动机加减速面积^[16],若节点能维持稳定,则滑差可逐渐恢复,若节点失稳,滑差将不断增大,可根据滑差的变化特性对节点稳定程度进行度量。而滑差可以通过在线辨识得到:基于辨识得到节点等值综合负荷模型参数后,若忽略感应电动机定子铜耗,可通过提取电动机有功来近似计算电磁转矩,利用转子运动方程积分即可得到转速,进而得到滑差^[10]。设故障发生时刻为 t_0 ,此时电动机滑差为 s_0 ,故障切除时刻为 t_c ,此时电动机滑差为 s_c ,首先通过转矩平衡的判据^[9-10]判断节点是否稳定,然后进行节点稳定度量指标的计算:

1) 若节点电压稳定,滑差 s 从 s_c 开始,在继续变化的过程中,设第一次恢复到 s_0 的时刻为 t_1 ,则定义滑差恢复时间为

$$K_r = \begin{cases} t_1 - t_c, & (t_1 - t_c \leq KT_r) \\ KT_r, & (t_1 - t_c > KT_r) \end{cases} \quad (1)$$

式中: T_r 为稳定节点监测时间窗; K 为临界状态的保守系数(下同)。经计算验证,结合在线评估的快速性及可靠性要求, T_r 及 K 的合适取值范围分别为0.5~1.2 s, 0.95~1。这里取 $T_r = 1$ s, $K = 0.95$ 。经最大-最小规范化处理,节点稳定度量指标为

$$K_{VS} = (1 - \frac{K_r - T_{\min}}{T_r - T_{\min}})(K_{VS\max} - K_{VS\text{mid}}) + K_{VS\text{mid}} \quad (2)$$

式中, $T_{\min} = 0$, $K_{VS\text{mid}} = 0$, $K_{VS\max} = 1$ 。

2) 若节点电压不稳定,设滑差 s 从 s_c 开始,增量达到 $(s_c - s_0)$ 的时刻为 t_1 ,此时滑差为 s_1 , s_c 为 s_0 和 s_1 的中位点,则定义滑差增长时间为

$$K_g = \begin{cases} t_1 - t_c, & (t_1 - t_c \leq KT_g) \\ KT_g, & (t_1 - t_c > KT_g) \end{cases} \quad (3)$$

式中 T_g 为不稳定节点监测时间窗口, T_g 的合适取值范围与 T_r 一致。取 $T_g = 1$ s, $K = 0.95$ 。经最大-最小规范化处理,节点稳定度量指标为

$$K_{VS} = \frac{K_g - T_{\min}}{T_g - T_{\min}}(K_{VS\text{mid}} - K_{VS\text{min}}) + K_{VS\text{min}} \quad (4)$$

式中, $T_{\min} = 0$, $K_{VS\text{mid}} = 0$, $K_{VS\text{min}} = -1$ 。

通过式(2)(4), K_{VS} 把节点稳定程度量化为-1~1的实数,-1表示极端失稳,1表示极端稳定,0表示临界稳定。

2.2 区域网络电压无功灵敏度矩阵

区域网络内节点之间的相互影响可用电压无功灵敏度矩阵进行描述。假定暂态过程中某一小段时间 Δt 内区域网络中各节点无功对电压的影响近似用线性关系描述, 即认为 Δt 内无功对电压的灵敏度 $\partial U/\partial Q$ 近似保持不变, 对 Δt 内在线 n 次观测得到的网络数据进行多元线性回归分析, 灵敏度即为回归方程中的回归系数。基本数学模型描述如下:

设区域网络中有 k 个节点, 各节点无功与节点 i 电压的灵敏度关系为

$$\Delta U_i = \sum_{j=1}^k \frac{\partial U_i}{\partial Q_j} \Delta Q_j + \varepsilon \quad (5)$$

式中 ε 表示近似描述带来的误差项。

多元线性回归模型如下:

$$\begin{cases} Y = X\beta + \varepsilon \\ E(\varepsilon) = 0, \text{COV}(\varepsilon, \varepsilon) = \sigma^2 I_n \end{cases} \quad (6)$$

该模型可简记为 $(Y, X\beta, \sigma^2 I_n)$, 其中 $Y =$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

I_n 为 n 阶单位矩阵。

针对单次观测有

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon \quad (7)$$

前后 2 次观测点的回归方程相减可得

$$\Delta y = \beta_0 + \beta_1 \Delta x_1 + \dots + \beta_k \Delta x_k + \Delta \varepsilon \quad (8)$$

由于 Δt 内认为灵敏度近似保持不变, 比较式 (5)~(8) 知, 式 (5) 中灵敏度的计算转化为式 (6) 中回归系数 $\beta_1 \sim \beta_k$ 的求解。暂态过程中短时间内参数变化较大, 可能导致模型预测不甚精确, 但是这种情况下的响应变化明显, 在较短响应时间内做分析时可靠性强^[17], 可表征节点间相互影响的强弱关系。为增强参数辨识结果的稳健性, 可适当加长观测时间, 通过多个观测点的平均化效应来减小偏差。结合前述节点监测时间窗 T_F/T_g 的设置及辨识计算速度的要求, 观测时间长度宜取为 0.5~1 s。为消除各节点无功变量间可能存在的有害共线性关系, 进一步通过逐步回归方法, 把可能存在共线性关系的变量中相对较弱的项及某些影响不显著的节点剔除出去。

对区域网络内的 k 个节点进行回归分析后, 可得灵敏度矩阵如下:

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2k} \\ \vdots & \vdots & & \vdots \\ \beta_{k1} & \beta_{k2} & \dots & \beta_{kk} \end{bmatrix} \quad (9)$$

式中 β_{ij} 表示第 j 个节点无功对第 i 个节点电压的灵敏度。对于矩阵的行向量 $\beta_i (i = 1, 2, \dots, k)$, 在 $\beta_{i1} \sim \beta_{ik}$ 之中, 设最大值为 $\beta_{i\max}$, 最小值为 $\beta_{i\min}$, 为避免不同度量单位对数据分析的影响, 对灵敏度系数进行规范化处理:

$$\alpha_{ij} = \frac{\beta_{ij} - \beta_{i\min}}{\beta_{i\max} - \beta_{i\min}} (\alpha_{\max} - \alpha_{\min}) + \alpha_{\min} \quad (10)$$

式中, $\alpha_{\min} = -1$, $\alpha_{\max} = 1$ 。若 $0 < \alpha_{ij} \leq 1$, 说明节点 j 有助于节点 i 电压稳定, 若 $-1 \leq \alpha_{ij} < 0$, 说明节点 j 不利于节点 i 电压稳定; 若 $\alpha_{ij} = 1$, 则表示区域网络内节点 j 最有助于维持节点 i 电压稳定, 若 $\alpha_{ij} = -1$, 则表示节点 j 最不利于维持节点 i 电压稳定, 若 $\alpha_{ij} = 0$, 则表示节点 j 对节点 i 电压稳定的影响可以忽略。

3 基于约束的半监督聚类

样本集的可靠划分, 是建立准确分类模型的基础。但电网中各节点多大范围、多大程度的失稳可以判断为区域性失稳, 目前并无可靠的量化标准或依据, 这给样本数据类标号的获取带来了较大困难。工程判据阈值法的固有缺点, 使其无法可靠地对样本做分类界定。若利用专家经验对样本逐个标记, 代价较大, 周期较长, 难以满足在线应用需求。若仅使用无标记数据进行无监督学习, 搜索过程带有一定的盲目性, 精度难以保证, 结果可能会与实际情况差距较大。结合已有认知, 在样本集内: 若样本所有节点都失稳, 显然这时区域稳定结果判为失稳; 若样本所有节点都能维持稳定且与临界稳定状态尚有一定裕度, 则区域稳定结果判为稳定。若把它们作为先验知识, 采用半监督聚类方法对数据集进行聚类分析, 既不用付出过大代价, 又能确保学习结果与实际相符^[18]。

半监督聚类中, **must-link** 和 **cannot-link** 为 2 种常见的成对约束关系, 分别表示 2 个样本必须划分到同一类、不能划分到同一类。类标号信息可转化为成对约束信息, 同为稳定类别或同为失稳类别的样本为 **must-link** 约束关系, 稳定类别样本与失稳类别样本间为 **cannot-link** 约束关系。采用基于硬性约束的半监督 K-means 算法 (constraint-partitioning K-means, COP-K-means)^[18] 进行聚类学习: 首先利用已标记的部分样本数据进行均值计算, 以此初始化聚类中心, 然后采用类标号的 **must-link** 和 **cannot-link** 约束关系作为监督信息进行 K-means 迭代搜索, 直至数据集的分簇稳定为止, 由此获得所有样本的类标号。

对于聚类结果质量的好坏, 利用内部评价法中

的轮廓系数指标^[18]进行评估。设数据集被划分为 k 个簇 C_1, C_2, \dots, C_k , 某个样本 i 属于第 j 个簇 C_j , 与同簇之内其他样本的平均距离(欧氏距离)为 a_i , 与簇 C_m 内样本的平均距离为 b_{jm} , $b_i = \min\{b_{jm} \mid m \neq j\}$ 为 i 与其他簇的簇内样本平均距离最小值。 a_i 反映 i 所属的簇的紧密度, 该值越小, 簇越紧凑; b_i 反映 i 与其它簇的分离度, 该值越大, i 与其它簇分离程度越大。样本 i 的轮廓系数为

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (11)$$

对 n 个样本的轮廓系数取平均值可得数据集整体的轮廓系数为

$$s = \frac{1}{n} \sum_{i=1}^n s_i \quad (12)$$

s 是整个数据集聚类结果紧密度和分离度的综合衡量指标, 以二维数据集为例(特征属性为 A_1 、 A_2 , 经规范化处理后二者取值范围均为 $0 \sim 1$), 如图 1 所示, 2 类散点分别代表 2 簇样本, s 越大, 聚类质量越好。

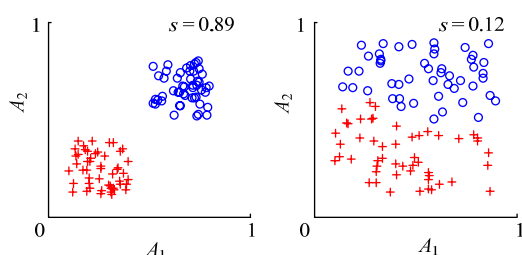


图 1 轮廓系数比较示意图

Fig. 1 Comparison of silhouette coefficients

4 评估实施流程

整个评估方案的实施流程如图 2 所示, 主要包含以下步骤:

1) 数据集集成及预处理。结合历史运行状况和可能事故, 通过仿真计算建立大量的仿真事件, 从

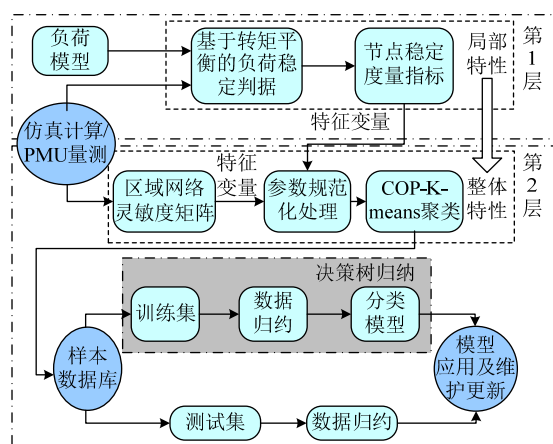


图 2 区域暂态电压稳定评估实施流程

Fig. 2 Flowchart of regional transient voltage stability assessment

中提取节点稳定度量指标、电压无功灵敏度矩阵等特征变量, 对数据进行规范化处理, 建立原始数据集, 利用 COP-K-means 算法对数据集进行有效划分, 建立样本数据库。为提高数据挖掘效率, 需进行数据归约, 采用基于相关性的特征选择算法(correlation-based feature selection, CFS)进行特征选择^[19], 以降低特征空间维数。

2) 基于决策树的分类训练。半监督聚类是实施决策树归纳的前提, 而后者可对前者的学习结果进行进一步的扩展挖掘和分析。本质上半监督聚类方法已能给出相对可靠的类别划分, 但其结果可解释性不强, 难以满足在线应用及进一步挖掘分析的要求, 为此, 继续对数据集进行训练学习, 利用决策树算法建立分类评估模型。在决策树算法中, 需选择某种特征属性度量标准作为分裂准则, 用以依次选取树中各节点特征属性及其最佳分裂点(量化分叉值), 从而产生纯度最高的分类结果。常见的度量标准有信息增益、信息增益率等^[18]。在某分裂点下, 某个特征属性相对于未分裂前可获得一定程度的有助于准确分类的信息增量, 以特征属性取值范围为上下限, 在所有可能的分裂点下, 信息增量最大者即为该属性的最佳分裂点, 相应的信息增量为其信息增益, 而信息增益率即为信息增益经规范化处理后的值。本文采用 C4.5 算法进行二叉决策树构建, 该算法以信息增益率为度量标准。在决策树构建过程中, 以训练集为对象, 首先计算特征空间内各特征属性的信息增益率, 取其中最大者作为树的根节点, 并依据最佳分裂点将数据集分为 2 个子集, 子集按照相同的方法从特征空间中继续选取特征属性, 使树继续生长, 相应的最佳分裂点使节点长出更多分枝。当某一分枝中的样本均属同一类时, 则生成终端节点(叶子节点), 当所有分枝都生成叶子节点时, 即完成树的构造。为防止树的过拟合, 使用后剪枝方法进行剪枝处理, 修剪样本数相对较少、对分类精度贡献不大的子树, 用子树中占多数的类来构造终端节点并进行替换, 使模型具有更强的泛化能力。

3) 模型评估及应用。在训练集中完成决策树分类学习后, 利用测试集对分类模型进行测试。若模型满足精度要求, 进一步提取出蕴藏在分类模型中的知识, 在决策树中所出现的特征属性即为关键特征属性(critical attribute), 所对应的分裂方式为关键分裂准则(critical splitting rule, CSR), 其中包含的关键节点(critical bus, CB)、节点间关键灵敏度系数(critical sensitivity coefficient, CSC)及相应的

CSR 构成区域暂态稳定的判据。结合具体的模型, 可进一步挖掘区域暂态电压失稳的内在规律。

4) 模型维护及更新。由于电力系统中的不确定性, 最初的样本集仅代表所能预测到的系统可能状况, 初始模型很难做到涵盖所有情况。为进一步提高模型的适应性及在线应用的可靠性, 对决策树模型进行逐步维护更新。在线监测过程中, 当出现检修、调整运行方式等情况后, 结合系统出现的变化, 生成新样本, 并添加到数据库中, 对决策树模型进行维护更新。在新模型指导下, 获取最新的 CB、CSC、CSR。当系统发生故障时, 即可通过暂态过程中辨识得到的 CB、CSC 属性值及最新的 CSR 对区域稳定性做出快速判断。故障时的系统运行状况, 同样可作为样本生成条件, 生成新样本, 对决策树模型进行更新。

5 算例分析及讨论

如图3所示, 以EPRI36节点系统为例, 把Bus16、Bus18、Bus19、Bus20、Bus21、Bus29组成的区域作为主要研究区域, 6个节点依次标号1—6。在系统正常运行的基础上, 调整负荷大小和发电机出力, 在潮流收敛的前提下, 按系统负荷的80%、100%、120%设定3种运行方式。动负荷占综合负荷的比例随机设为20%、50%、100%。暂态稳定计算中, 在区域中各线路首端的0、20%、40%、60%、80%等处随机设置故障位置; 设某条线路的某个位置在0.1 s时发生三相短路故障, 0.15 s切除近端故障, 0.20 s切除远端故障, 考虑后备保护动作时, 0.45 s切除近端故障, 0.5 s切除远端故障^[20]。考虑断线后网架结构变化的情况, 分别设置 Bus16-Bus19、Bus19-Bus21 断线, 在 Bus16-Bus18 和 Bus16-Bus20 不同位置设置短路故障。以仿真计算数据来模拟 PMU 量测数据。在上述条件设定下, 通过 PSASP 仿真生成 900 个样本。

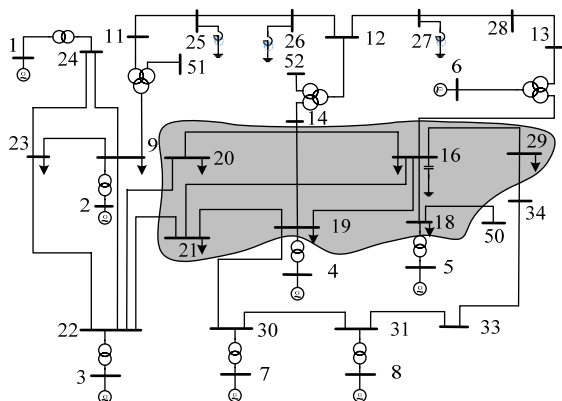


图3 EPRI 36 节点系统
Fig. 3 EPRI 36-bus system

1) 灵敏度辨识验证。

从样本集中随机抽取样本, 首先进行灵敏度辨识的验证。以 Bus16 的灵敏度辨识为例, 观测时间长度分别取为 0.1 s 和 0.5 s, 根据回归模型, 利用测量得到的无功变化信息对 Bus16 的电压进行预测, 以故障切除时刻为 0 时刻, 对故障切除后的电压进行预测, 预测曲线如图 4 所示。

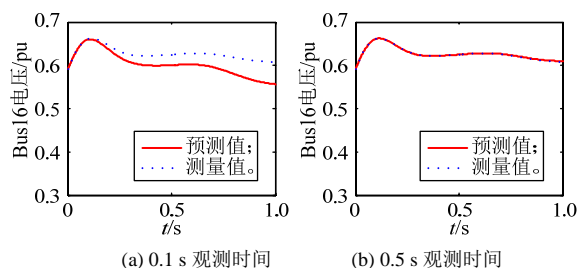


图4 回归模型预测曲线

Fig. 4 Prediction curves of the regression model

由图可知, 观测时间为 0.1 s 时, 预测精度仅在 0.15 s 内相对较高, 随后不断降低, 误差不断增大, 说明 0.1 s 内辨识得到的灵敏度难以准确反映随后的响应特性。而观测时间增加到 0.5 s 后, 在随后的 0.5 s 内均能保持较高精度, 预测曲线与辨识曲线基本重合, 说明辨识得到的灵敏度已具备较强的稳健性, 能可靠反映 1 s 时间尺度内系统的响应特性。因此, 为更加可靠地提取特征变量, 在对节点间电压无功灵敏度进行辨识的过程中, 观测时间长度均取为 0.5 s。

2) 半监督聚类有效性评估。

对样本集进行 COP-K-means 半监督聚类分析。为进行比较, 按照工程阈值法对样本集进行分类: 当区域内最低节点电压低于 0.75 pu 的时间超过 1 s, 则判为区域失稳, 否则判为稳定。把工程阈值法相应的类内样本均值等效为簇中心, 对 2 种方法得到的分类结果比较, 如表 1 所示。COP-K-means 算法与工程阈值法的分类结果相比, 聚类质量明显优于后者, 说明半监督聚类算法对数据集的划分更为可靠、合理。

表1 聚类有效性评估比较
Tab. 1 Comparison of clustering validity

评估指标	工程阈值法	COP-K-means
轮廓系数	0.136 8	0.307 2

3) 模型评估及深入挖掘。

通过 COP-K-means 聚类学习后, 即可获得所有样本的类标号。原始特征空间为包含 6 个节点稳定度量指标和 36 个灵敏度系数的 42 维空间, 经 CFS 特征选择, 降维后特征子集为 $\{K_{VS_Bus16}, K_{VS_Bus18}, K_{VS_Bus19}, K_{VS_Bus21}, K_{VS_Bus29}, \alpha_{22}, \alpha_{23}, \alpha_{24}, \alpha_{52}, \alpha_{54}\}$ 。采用

C4.5 决策树归纳算法对数据集进行分类训练, 训练集与测试集的划分采用 10-折交叉验证方式。首先将样本集随机分为 10 个子集, 每个子集样本数均为 90, 训练和测试均进行 10 次, 在第 i ($1 \leq i \leq 10$) 次迭代中, 第 i 个子集作为测试集, 其他子集作为训练集, 按照前述决策树的构造方法进行分类学习。迭代完后, 性能最优的决策树模型选为最终的分类模型, 如图 5 所示。

图 5 中终端节点的标号 1 代表稳定类别, 标号 0 代表失稳类别。10-折交叉验证下, 为减小分类评估的偏倚, 以各次迭代测试中测试集分类准确率的平均值作为模型最终分类准确率。由此得到的分类准确率为 98.9%, 说明模型分类性能较好。在决策树模型中, Bus18、Bus19、Bus21、Bus29 作为关键节点, Bus18 对 Bus21 的灵敏度 α_{24} 作为节点间关键灵敏度系数, 这些关键特征属性及树中相应的量化分叉规则构成关键分裂准则, 即为区域暂态电压稳定判断。

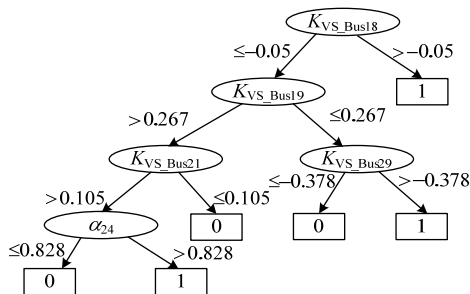


图 5 初始数据集训练得到的决策树模型
Fig. 5 Decision tree model for initial data sets

结合模型和样本数据, 进一步挖掘特征变量间的关联特性和区域暂态电压稳定性的内在规律。半监督聚类分析得到的 2 聚类中心(分别用 C_1 、 C_2 表示)代表样本空间各特征变量的均值。分析节点稳定度量指标 K_{VS_Bus16} — K_{VS_Bus29} 的均值, 如表 2 所示, K_{VS_Bus16} 、 K_{VS_Bus18} 、 K_{VS_Bus29} 的均值十分接近, 且相对较小; K_{VS_Bus19} 、 K_{VS_Bus21} 的均值也很接近, 且相对较大。由此可得, Bus16、Bus18、Bus29 具有一致的稳定性, 且维持稳定能力较弱, Bus19 和 Bus21 也具有一致的稳定性, 且维持稳定能力较强, 而 Bus20 维持稳定能力相对居中。据此将各节点分为 3 个子区: $A_1=\{\text{Bus16, Bus18, Bus29}\}$, $A_2=\{\text{Bus20}\}$, $A_3=\{\text{Bus19, Bus21}\}$, A_1 为最易失稳

表 2 特征变量均值比较						
Tab. 2 Comparison of feature variables' mean values						
聚类中心	Bus16	Bus18	Bus19	Bus20	Bus21	Bus29
C_1	0.642 1	0.625 6	0.587 8	0.526 7	0.568 6	0.641 8
C_2	-0.558 8	-0.575 4	-0.394 5	-0.329 6	-0.387 9	-0.544 2

子区, A_3 为最难失稳子区。根据分区结果, 结合决策树中的关键节点, 将节点 Bus18 作为 A_1 的代表节点, Bus19 作为 A_3 的代表节点, 各子区内节点的相互影响及子区之间的相互影响可通过相应灵敏度系数度量。进一步分析决策树模型, 当 A_1 未发生较严重失稳时, 区域可维持稳定; 当 A_1 发生较严重失稳而 A_3 也不具备较强稳定性时, 区域容易发生失稳; 当 A_2 对 A_3 具有较强支撑作用, 可在一定程度上增强区域稳定性。

不失一般性地从样本集中随机抽样, 抽到的样本事件暂态过程中各节点电压变化曲线如图 6 所示。图 6 中各节点电压变化趋势与上述分区及代表节点选取是相符的, 由此进一步证明上述挖掘分析的合理性。因此, 建立的评估模型可为电压分区和分区代表节点的选取提供数据分析层面的依据, 为在线监测时区域内关键观测点的选取提供指导。

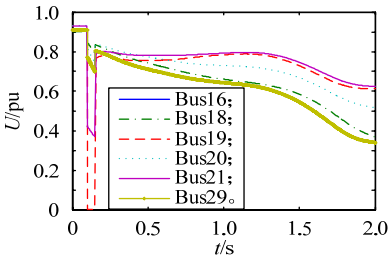


图 6 节点电压变化曲线
Fig. 6 Variation curves of different buses' voltages
4) 模型维护更新。

考虑新的暂态事故, 同样在 Bus16-Bus19 或 Bus19-Bus21 发生断线, 而发生三相短路的线路为 Bus16-Bus21、Bus16-Bus29, 生成 360 个初始训练过程中未曾出现的新样本。模型维护更新前, 首先利用新样本测试初始模型的适应性。360 个样本中, 稳定类别判断正确的数目为 349, 误判数目为 4, 漏判数目为 7, 判别准确率为 96.9%。由此可知, 维护更新前, 模型已具有较强适应性。利用新样本对模型进行更新训练, 更新后的决策树模型如图 7 所示, 分类准确率达 98.6%。对比图 5 和图 7, 新模型基本结构变化不大, 仅是关键属性及关键分裂

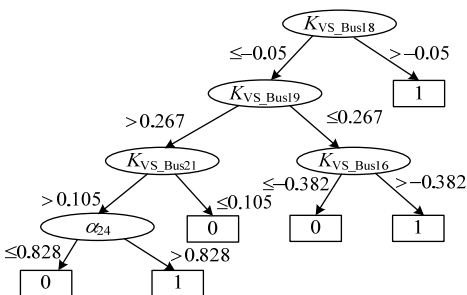


图 7 更新后的决策树模型
Fig. 7 Updated decision tree model

准则略有变化。可以预见, 面对新的未知事件时, 更新后模型将具有更强的适应性和更高的准确性。

6 结语

本文详细阐述了基于数据挖掘的区域暂态电压稳定评估方案。建立了两层分析框架, 充分利用暂态过程中局部节点稳定性的已知规律来评估未知的区域整体稳定性。通过实测辨识得到电压无功灵敏度矩阵, 以此作为反映区域网络节点间相互影响的重要特征变量。采用半监督学习方法对数据集进行有效聚类, 克服了区域性失稳界定尚无可靠标准的困难。利用决策树算法获得了基于逐步更新分类模型的区域暂态电压稳定判据。仿真算例验证了评估方案的有效性, 并通过分类模型挖掘出电压分区、代表节点等选取依据及各分区稳定性与区域稳定性之间的内在规律。

值得注意的是, PMU 在线量测的可能误差和不确定性尚未考虑在内, 考虑量测误差的分类评估及分类模型的优化还需进一步完善。

参考文献

- [1] 陈亦平, 洪军. 巴西“11.10”大停电原因分析及对我国南方电网的启示[J]. 电网技术, 2010, 34(5): 77-82.
Chen Yiping, Hong Jun. Analysis on causes of blackout occurred in Brazilian power grid on Nov.11, 2009 and lessons drawn from it to China Southern power grid[J]. Power System Technology, 2010, 34(5): 77-82(in Chinese).
- [2] 毛安家, 张戈力, 吕跃春, 等. 2011年9月8日美墨大停电事故的分析及其对我国电力调度运行管理的启示[J]. 电网技术, 2012, 36(4): 74-78.
Mao Anjia, Zhang Geli, Lü Yuechun, et al. Analysis on large-scale blackout occurred in south America and north Mexico interconnected power grid on Sept. 8, 2011 and lessons for electric power dispatching in China[J]. Power System Technology, 2012, 36(4): 74-78(in Chinese).
- [3] 梁志峰, 葛睿, 董昱, 等. 印度7.30、7.31大停电事故分析及对我国电网调度运行工作的启示[J]. 电网技术, 2013, 37(7): 1841-1848.
Liang Zhifeng, Ge Rui, Dong Yu, et al. Analysis of large-scale blackout occurred on July 30 and July 31, 2012 in India and its lessons to China's Power Grid dispatch and operation[J]. Power System Technology, 2013, 37(7): 1841-1848(in Chinese).
- [4] 汤涌. 电力系统电压稳定性分析[M]. 北京: 科学出版社, 2011: 10.
- [5] 吴浩, 郭瑞鹏, 韩祯祥. 电力系统微分代数模型的奇异性和暂态电压稳定[J]. 电力系统自动化, 2006, 30(13): 16-21.
Wu Hao, Guo Ruipeng, Han Zhenxiang. Relationship between transient voltage stability and singularity in power system differential algebraic model[J]. Automation of Electric Power Systems, 2006, 30(13): 16-21(in Chinese).
- [6] 廖国栋, 谢欣涛, 侯益灵, 等. 湖南电网暂态电压稳定性问题[J]. 电网技术, 2012, 36(10): 196-202.
Liao Guoding, Xie Xintao, Hou Yiling, et al. Study on transient voltage stability of Hunan power grid[J]. Power System Technology, 2012, 36(10): 196-202(in Chinese).
- [7] 杨银国, 林舜江, 欧阳逸风, 等. 三级电压控制体系下大电网暂态电压安全仿真及控制策略[J]. 电网技术, 2012, 37(4): 1045-1051.
Yang Yinguo, Lin Shunjiang, Ouyang Yifeng, et al. Transient voltage security simulation and control strategy in large-scale power grids under three-level voltage control[J]. Power System Technology, 2012, 37(4): 1045-1051(in Chinese).
- [8] 汪娟娟, 张尧, 夏成军, 等. 交直流电力系统暂态电压稳定性综述[J]. 电网技术, 2008, 32(12): 30-34.
Wang Juanjuan, Zhang Yao, Xia Chengjun, et al. Survey of studies on transient voltage stability of AC/DC power system[J]. Power System Technology, 2008, 32(12): 30-34(in Chinese).
- [9] 孙华东. 计及感应电动机负荷的电压稳定分析及应用研究[D]. 北京: 中国电力科学研究院, 2005.
- [10] 李立理. 综合负荷的主导参数辨识及暂态电压稳定评估与监测[D]. 北京: 清华大学, 2009.
- [11] Senroy N, Heydt G T, Vittal V. Decision tree assisted controlled islanding[J]. IEEE Transactions on Power Systems, 2006, 21(4): 1790-1797.
- [12] Sun K, Likhate S, Vittal V, et al. An online dynamic security assessment scheme using phasor measurements and decision trees[J]. IEEE Transactions on Power Systems, 2007, 22(4): 1935-1943.
- [13] 姚德全, 贾宏杰, 赵帅. 基于复合神经网络的电力系统暂态稳定评估和裕度预测[J]. 电力系统自动化, 2013, 37(20): 41-46.
Yao Dequan, Jia Hongjie, Zhao Shuai. Power system transient stability assessment and stability margin prediction based on compound neural network[J]. Automation of Electric Power Systems, 2013, 37(20): 41-46(in Chinese).
- [14] 叶圣永, 王晓茹, 刘志刚, 等. 基于支持向量机增量学习的电力系统暂态稳定评估[J]. 电力系统自动化, 2011, 35(11): 15-19.
Ye Shengyong, Wang Xiaoru, Liu Zhigang, et al. Power system transient stability assessment based on support vector machine incremental learning method[J]. Automation of Electric Power Systems, 2011, 35(11): 15-19(in Chinese).
- [15] Carson W Taylor. Power system voltage stability[M]. New York, United States: McGraw-Hill, 1994.
- [16] 黄志刚. 暂态电压稳定实用判据研究[D]. 北京: 清华大学, 2007.
- [17] 张俊勃. 基于广域信息的电力系统动态建模与应用[D]. 北京: 清华大学, 2013.
- [18] Han J, Kamber M, Pei J. Data mining: concepts and techniques(Third edition)[M]. Burlington, United States: Morgan kaufmann, 2011.
- [19] Hall M A. Correlation-based feature selection for machine learning[D]. Hamilton, New Zealand: The University of Waikato, 1999.
- [20] 叶圣永, 王晓茹, 刘志刚, 等. 基于支持向量机的暂态稳定评估双阶段特征选择[J]. 中国电机工程学报, 2010, 30(31): 28-34.
Ye Shengyong, Wang Xiaoru, Liu Zhigang, et al. Dual-stage feature selection for transient stability assessment based on support vector machine[J]. Proceedings of the CSEE, 2010, 30(31): 28-34(in Chinese).



朱利鹏

收稿日期: 2014-09-26。

作者简介:

朱利鹏(1990), 男, 硕士研究生, 主要研究方向为电力系统运行与控制, E-mail: zhulpwhu@126.com;

陆超(1977), 男, 博士, 副教授, 博士生导师, 主要研究方向为电力系统稳定分析与控制;

孙元章(1953), 男, 教授, 博士生导师, 主要研究方向为电力系统分析与规划、大规模风电接入运行与控制技术、广域测量技术的高级应用。

(责任编辑 李兰欣)