

基于仿射传播聚类算法的广义负荷稳态特性建模及其应用

褚壮壮¹, 梁 军¹, 张 旭², 董晓明¹, 张永亮³

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

2. 国网济南市供电公司, 山东 济南 250012; 3. 国网滨州市供电公司, 山东 滨州 256600)

摘要: 大规模可再生新能源并网增加了网络节点功率流向的不确定性, 使节点的源、荷特性不再清晰, 给广义负荷建模带来了新的挑战。因此提出一种基于仿射传播(AP)聚类算法的广义负荷稳态特性聚类方法。对广义负荷节点根母线功率数据进行特征分析, 利用动力学的波动强度理论选取功率波动序列的最小时间长度。以时段内各最小时间长度的样本波动强度序列以及时段内样本数据数字特征为指标构造日时段特征向量, 并以该特征向量为聚类指标, 应用 AP 聚类算法自适应调整建模数据的聚类数和聚类中心。通过引入概率信息的广义负荷建模方法对各聚类类别样本建模并检验聚类方法的有效性。算例分析表明, 该聚类方法自适应确定聚类类别且聚类效果好, 能够充分反映日时段特性, 聚类后通过特性综合得到精确广义负荷模型, 应用于风电接入后的风险分析仿真计算。

关键词: 负荷建模; 负荷特性; 仿射传播聚类算法; 波动强度; 时段特性; 广义负荷特性; 聚类; RBF 神经网络
中图分类号: TM 743 **文献标识码:** A **DOI:** 10.16081/j.issn.1006-6047.2016.03.018

0 引言

大规模新能源并网对电力系统安全稳定运行造成很大冲击, 可再生能源发电的接入改变了电力网络节点的功率传输走向及其特性, 特别是分布式新能源的不断涌现, 使传统的负荷节点向电网倒送功率成为可能^[1]。因此, 含新能源的广义负荷建模是对传统负荷建模在新场景下的发展和延伸。新能源具有随机波动性、间歇性, 而负荷本身具有时变性, 这两者的相互作用加剧了广义负荷节点的不确定性, 对系统潮流分布、仿真计算、电网安全运行等会产生较大影响^[2-5], 因此考虑风电随机波动性的广义负荷节点特性建模对电力系统分析具有重要意义。

传统建模方法采用分类与综合的思想, 通过总体测辨法得到负荷的准确模型, 拟合效果较好^[6-13], 推进了负荷建模工作的发展和应用。其中文献[6]首次提出了动态负荷特性的分类与综合问题, 并提出了基于系统回响辨识和基于模型回响辨识的动态负荷特性综合办法; 文献[7]以标准电压激励响应和负荷有功运行水平作为特征向量, 采用 KOHONEN 神经网络映射到高维空间进行分类; 文献[8]对 K-means 聚类算法进行改进并定义了评估指标, 以此进行负荷聚类分析; 文献[9]运用模糊聚类方法给出了长沙主要变电站的分类结果; 文献[10]通过日负荷最大、最小负荷率构造分类判据, 将特殊负荷分为电解铝类负

荷和铁合金类负荷, 并分别对 2 类特殊负荷模型进行验证; 文献[11]以感应电动机综合负荷模型参数和动态负荷所占比例为特征向量, 提出了基于模糊 C 均值聚类的负荷特性分类方法, 采用直接综合和加权平均综合 2 种综合方法对分类结果进行综合建模; 文献[12]以各负荷节点-变电站的不同类型负荷比例为特征向量, 运用基于模糊等价关系的传递闭包法对实测数据进行了模糊分类; 文献[13]建立山峰密度函数自适应确定聚类数和聚类中心。

传统负荷建模聚类方法处理负荷的时变性效果较好, 但其采用的聚类方法需人为设定聚类数、聚类中心等, 较为主观, 在考虑风电接入的复杂场景下不具有普遍适用性。由于风电接入规模的不断扩大, 渗透率逐渐增加, 改变了传统负荷节点单纯消耗功率的情况, 使得负荷节点组成成分和功率流向发生本质改变, 传统简单聚类策略及聚类方法对新场景下由风电和负荷叠加产生的复杂数据进行合理聚类存在不足, 因此研究考虑风电接入新场景下广义负荷不确定性的合理聚类与综合以及建模成为亟待解决的问题。

为解决上述问题, 本文提出基于仿射传播 AP (Affinity Propagation) 聚类算法的广义负荷稳态特性的聚类与综合方法。该聚类算法属于无监督聚类算法, 因特殊的消息传递机制与竞争机制自适应确定聚类数, 能够自动消除聚类振荡并自动寻找最优聚类结果, 面对复杂场景, 聚类质量高, 效果明显。首先对节点根母线功率数据进行特征分析, 利用动力学的波动强度理论选取功率波动序列的最小时间长度, 然后以时段序列内各最小时间长度的波动强度

以及时段序列的数字特征为指标构造特征向量,利用日时段特征向量作为聚类指标,应用 AP 聚类算法自适应调整样本数据的聚类数和聚类中心。采用依概率分区间的广义负荷建模方法^[14],利用 RBF 神经网络对各类样本数据建立概率广义负荷模型,并通过模式验证判断测试样本所属类别,以确定待测试样本属类并检验该聚类方法的有效性。通过仿真算例与 K -means 算法对比,说明 AP 聚类算法聚类质量优越,解决了 K -means 聚类随机初始聚类中心导致的聚类结果不稳定问题。最后将聚类后的模型应用于风电接入后的风险分析仿真计算,验证了本文方法的实用性,可为系统决策提供参考。

1 AP 聚类算法描述

传统负荷建模领域通常采用 K -means 聚类算法^[15-16]、模糊神经网络聚类算法^[7-8,11]等,大部分需人为设定聚类数和聚类中心等,主观因素较强,不具备客观普遍性,因此本文引入 AP 聚类算法^[17]。AP 聚类算法是一种新的无监督聚类算法,无需事先定义类数。算法开始时把所有的数据点均视作类中心,通过数据点间的“信息传递”来实现聚类过程。在迭代过程中不断搜索合适的聚类中心,自动从数据点间识别聚类中心的位置及个数。AP 算法是在数据点的相似度矩阵上进行聚类的,聚类的目标是使数据点与其聚类中心之间的距离达到最小,因此选用欧氏距离作为相似度的测量指标,即任意 2 个向量 X_i 和 X_k 的相似度为:

$$S(i, k) = -d^2(X_i, X_k) = -\|X_i - X_k\|_2 \quad (1)$$

在聚类之前,每个数据点被赋予偏向参数 $P(i) = S(i, i)$,表示数据点 i 被选作聚类中心的倾向性,该值越大则聚类数越多。AP 聚类算法用代表矩阵 $[r(i, k)]$ (responsibility) 和 适应矩阵 $[a(i, k)]$ (availability) 来表示数据点之间的 2 类信息,其中 $r(i, k)$ 是从 X_i 指向候选聚类中心 X_k ,它反映了 X_k 适合作为 X_i 的聚类中心所积累的证据,该值越大表明候选聚类中心 k 成为真正聚类中心的可能性越大; $a(i, k)$ 从候选聚类中心 X_k 指向 X_i ,它反映了 X_i 选择 X_k 作为其聚类中心的合适程度所积累的证据,如图 1 所示。在邻近传

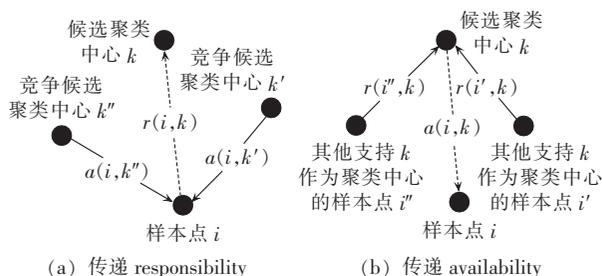


图 1 AP 算法消息传递机制

Fig.1 Message transfer mechanism of AP algorithm

播结束时, X_i 的聚类中心确定为 X_k , k 满足:

$$\arg \max_k (a(i, k) + r(i, k)) \quad (2)$$

基于 AP 算法的聚类过程具体计算步骤如下。

a. 初始化。确定样本点数 N 、迭代次数 M 、聚类样本数据的特征向量,按式(1)计算 N 个样本点的相似度矩阵 S 作为输入量,其中对偏向参数 P 赋值,初始化 r 和 a 为 0。

b. 计算各点间 $r(i, k)$ 和 $a(i, k)$ 。

$$r(i, k) \leftarrow S(i, k) - \max_{j \neq k} \{a(i, j) + S(i, j)\} \quad (3)$$

$$a(i, k) \leftarrow \min \left\{ 0, r(k, k) + \sum_{j \neq i \& j \neq k} \{ \max \{ 0, r(j, k) \} \} \right\} \quad (4)$$

$$r(k, k) \leftarrow P(k) - \max_{j \neq k} \{a(k, j) + S(k, j)\} \quad (5)$$

$$a(k, k) \leftarrow \sum_{j \neq k} \max \{ 0, r(j, k) \} \quad (6)$$

c. 由于 AP 聚类过程中容易产生振荡,因此通过设定阻尼系数 λ 来控制迭代速度,循环迭代更新 r 和 a ,如式(7)和(8)所示。

$$r_i = (1 - \lambda)r_i + \lambda r_{i-1} \quad (7)$$

$$a_i = (1 - \lambda)a_i + \lambda a_{i-1} \quad (8)$$

d. 迭代满足式(2)所确定的 k 点,则 k 为样本点 i 的聚类中心。如果迭代次数超过设定的最大值或者当聚类中心在若干次迭代中不发生改变时终止计算,确定类中心及各类的样本点;否则返回步骤 **b** 继续计算。

2 基于日时段功率空间的特征向量

在传统负荷聚类与综合研究中,负荷场景简单,通过简单的聚类方法以及聚类策略进行聚类就可将样本数据区分不同的类别,而大规模风电接入负荷侧后,由于负荷时变性与风功率波动性相互叠加,节点呈现的特性包括幅值和流向具有随机性,场景更为复杂,需提出新的指标作为评价标准。广义负荷节点有功功率因其能够可靠表征节点特性、变化范围大、易于细化分段的特点,不但可定性分析节点呈现负荷特性还是电源特性,还可以将其定量细化到具体功率范围,因此以节点有功功率作为节点特性特征参数,通过时段细化节点功率空间,以时段内各最小时间长度的样本波动强度序列以及时段内样本数据统计量为指标构造日时段特征向量进行聚类分析。

2.1 波动强度概念

为构建聚类所需的特征向量,引入场均速度、波动速度和波动强度的概念。波动强度 (fluctuation intensity) 是动力学领域的一种统计物理概念^[18],可以表征信号序列曲线的波动程度,定义为波动速度的均方根与场均速度的比值,某一段序列波动强度越小则信号波动越小,反之波动越剧烈。

波动强度数学表达式为:

$$\begin{cases} \bar{P} = \frac{1}{T_i} \int_{T_i} p(t) dt \\ P'(t) = p(t) - \bar{P} \\ \gamma = \frac{\sqrt{\frac{1}{T_i} \int_{T_i} (p(t) - \bar{P})^2 dt}}{\bar{P}} = \frac{\sqrt{\frac{1}{K} \sum_{i=1}^K (p(i) - \bar{P})^2}}{\bar{P}} \end{cases} \quad (9)$$

其中, γ 为波动强度; \bar{P} 为场均速度; $P'(t)$ 为波动速度; K 为样本序列的点数; T_i 为采样周期; $p(i)$ 为样本序列各点值。

2.2 波动强度最小时间长度

定义波动强度最小时间长度 T 为能保持该时段内广义负荷母线功率近似不变的最小时间段, 为此该值的选取应满足在 T 内波动强度变化率不超过规定的变化率阈值, 则近似认为该时段内近似功率不变, 可将本时间长度内功率波动值作为特征向量的元素; 同时为保证聚类数合适, 最小时间长度不能过小, 故 T 应取满足式(10)的最小值。

$$\begin{cases} T = \arg \min_t (\max_i \{\phi_{it}\} < \sigma) \quad t = N_t t_{\min} \left(N_t \in \mathbf{N}^+, N_t \leq \frac{t_{\max}}{t_{\min}} \right) \\ \phi_{it} = \sqrt{\frac{t_{\min}}{t} \sum_{j=\frac{t}{t_{\min}}(i-1)+1}^{\frac{t}{t_{\min}} i} (P\{j\} - \bar{P}_{av})^2} / \bar{P}_{av} \quad i=1, 2, \dots, \frac{t_{\max}}{t} \\ \bar{P}_{av} = \left(\sum_{k=\frac{t}{t_{\min}}(i-1)+1}^{\frac{t}{t_{\min}} i} P\{k\} \right) / \left(\frac{t}{t_{\min}} \right) \end{cases} \quad (10)$$

其中, t_{\min} 和 t_{\max} 分别为采样时间和样本总体时间; t 为待定时间长度, 取采样时间的倍数, 以分钟计; N_t 为不大于采样样本数的自然数; ϕ_{it} 为待定时间长度 t 内, 采样样本序列 i 的功率波动强度; T 为满足约束条件 $\max \{\phi_{it}\} < \sigma$ 的最小时间长度; P 为负荷与风电组成的广义负荷根母线有功功率序列; \bar{P}_{av} 为待定时间长度 t 下, 采样序列 i 的有功功率均值; j, k 为功率序列号; σ 为变化率阈值。

2.3 聚类特征向量

将全部训练样本数据按聚类时间间隔 T_j 统一分段。聚类时间间隔 T_j 由多个 T 组成, 利用 T_j 内每个最小时间长度 T 内样本的变化波动序列表征数据波动趋势, 并和实际有功功率统计量构成特征向量 W_p :

$$W_p = [\phi_{p1}, \phi_{p2}, \dots, \phi_{pb}, W_{\max p}, W_{\min p}, \bar{W}_p] \quad (11)$$

其中, p 为聚类时间间隔序列号; b 为 T_j 中包含最小时间长度 T 的序列数; $W_{\max p}$ 和 $W_{\min p}$ 分别为第 p 个聚类时间间隔内的最大有功功率和最小有功功率; \bar{W}_p 为第 p 个聚类时间间隔内有功功率的均值; $\phi_{p1}, \phi_{p2}, \dots, \phi_{pb}$ 为 T_j 内 b 个最小时间长度波动强度序列。以 ϕ_{pi} 为例, 其计算式如下:

$$\begin{cases} \phi_{pi} = \sqrt{\frac{t_{\min}}{T} \sum_{j=t_1}^{t_2} (P\{j\} - \bar{P}_{av})^2} / \bar{P}_{av} \\ \bar{P}_{av} = \left(\sum_{j=t_1}^{t_2} P\{j\} \right) / (T/t_{\min}) \\ t_1 = (i-1)T/t_{\min} + (p-1)T_j/T + 1 \\ t_2 = t_1 + T/t_{\min} - 1 \end{cases} \quad (12)$$

其中, i 为 T_j 中最小时间长度序列号; P 为负荷与风电组成的根母线有功功率序列; j 为广义负荷有功功率序列标号; \bar{P}_{av} 为最小时间长度序列 i 内功率均值。

由波动强度序列计算公式可知, 其幅值大小充分反映广义负荷节点功率波动的幅度, 其值正负反映广义负荷节点功率的流向, 因此以各最小时间长度内样本的波动强度序列为基础所构造的特征向量作为聚类指标可以合理区分不同时段广义负荷节点特性。

3 基于概率标识的广义负荷建模

通过构造日时段的特征向量并利用 AP 聚类算法可以获得复杂场景样本数据聚类结果, 但如何得到每类数据的广义负荷模型以验证聚类效果, 需采用广义负荷特性综合的方法。由于风电接入使得广义负荷节点的功率流向和幅值大小呈现不确定性, 传统建模方法不具备随机特征描述能力, 难以应用于风电接入不确定场景下的建模分析, 因此本文采用带有概率信息的广义负荷建模进行聚类验证。

3.1 节点特性提取

针对风电接入后节点功率流向的改变, 有功功率因其能够可靠表征节点特性、变化范围大而被选作节点特性参考变量。以消耗功率或发出功率为依据, 将节点特性划分为负荷特性或电源特性; 针对节点特性的不确定性变化, 对有功功率样本空间进行自适应分段并统计其概率分布; 对有功功率依据特性自适应划分区间, 利用 RBF 神经网络法学习并提取各区间节点特征。

RBF 神经网络是多维空间插值的传统技术, 可以经过每个样本点, 能够逼近任意的非线性函数, 处理难以解析的规律性, 具有良好的泛化能力和全局逼近能力, 并有很快的学习收敛速度, 克服了 BP 神经网络存在的局部最小值和收敛速度慢的缺陷^[19], 因此本文采用 RBF 神经网络作为节点特性提取的模型。模型结构由输入层、隐含层和输出层组成, 其网络结构如图 2 所示。

该网络从输入层到隐含层为非线性映射, 隐含层到输出层为线性映射。隐含层径向基函数通常选用高斯核函数。RBF 神经网络属于前向型神经网络, 其结构具有自适应性, 且输出与初始权值无关。相比其他前向型网络, RBF 网络具有结构简单、训练简洁、收敛速度快、逼近性能好、需设置参数少等特点, 因此被广泛应用于非线性优化、时间序列预测和模

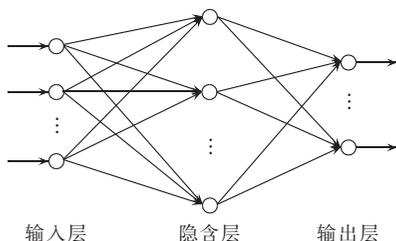


图 2 RBF 神经网络结构

Fig.2 Structure of RBF neural network

式识别等科学领域。

本文采用 RBF 神经网络函数对区间样本进行模型特征提取。其中,输出变量为有功功率 p ,输入变量为节点电压 u ,表达式如式(13)所示:

$$p(u) = \sum_{r=1}^M w_r R_r(u) \quad (13)$$

其中, w_r 为输出层连接权值; M 为区间样本个数; $R_r(u)$ 为第 r 层隐含层径向基函数。

计算模型结构如下:

$$\begin{cases} \min E_m = \frac{1}{2} \sum_{n=1}^{N_m} \sum_{k=1}^{N_o} e_{k,n}^2 \\ e_{k,n} = P_{k,n} - p_{k,n} \\ p_{k,n} = \sum_{j=1}^{N_h} w_{j,k} R(\mathbf{X}_n) \\ R(\mathbf{X}_n) = \exp\left(-\frac{\|\mathbf{X}_n - \mathbf{C}_j\|_2}{2\delta_j^2}\right) \\ \mathbf{X}_n = [u_{1n}, u_{2n}, \dots, u_{N_n}] \end{cases} \quad (14)$$

其中, k 为输出变量序号; n 为输入变量序号; m 为节点分段功率区间编号; E_m 为区间样本训练误差; N_m 为区间样本数; N_o 为输出神经元个数; $P_{k,n}$ 为区间样本功率实测值; $p_{k,n}$ 为模型计算值; \mathbf{X}_n 为输入变量向量; \mathbf{C}_j, δ_j 分别为第 j 个隐含层神经元的中心和扩展常数; N_h 为隐含层神经元的个数; $w_{j,k}$ 为第 j 个输出变量与第 k 个隐含层神经元的连接权值; N_i 为输入层神经元个数。

采用梯度自适应调整算法求解模型参数,调整公式为:

$$\begin{cases} \Delta C_j = \frac{\eta}{\delta_j^2} \sum_{n=1}^{N_m} \sum_{k=1}^{N_o} e_{k,n} w_{j,k} R(\mathbf{X}_n) (\mathbf{X}_n - \mathbf{C}_j) \\ \Delta \delta_j = \frac{\eta}{\delta_j^3} \sum_{n=1}^{N_m} \sum_{k=1}^{N_o} e_{k,n} w_{j,k} R(\mathbf{X}_n) \|\mathbf{X}_n - \mathbf{C}_j\|_2 \\ \Delta w_{j,k} = \eta \sum_{n=1}^{N_m} \sum_{k=1}^{N_o} e_{k,n} R(\mathbf{X}_n) \end{cases} \quad (15)$$

其中, ΔC_j 为误差对隐含层中心的调整参数; $\Delta \delta_j$ 为误差对隐含层扩展常数的调整参数; $\Delta w_{j,k}$ 为误差对隐含层输出权值的调整参数; η 为训练学习系数。

3.2 模型结构

每类合并各段模型,形成如式(16)的统一模型结构。

$$P(u) = \begin{cases} p_i^s(u_i^s) |_{P_s} & i=1, 2, \dots, i_{\max} \\ p_i^l(u_i^l) |_{P_l} & i=i_{\max}+1, \dots, i_{\max} \end{cases} \quad (16)$$

其中, $i_{\max} = 10P_{\min}/P_{\text{base}}$, $i_{\min} = 10P_{\max}/P_{\text{base}}$, P_{\min}, P_{\max} 分别为节点电源特性功率最小值和最大值, P_{base} 为基准功率(本文中取风场基准功率 100 MW); $p_i^s(u_i^s), p_i^l(u_i^l)$ 分别为各段下提取的电源特性和负荷特性关系表达式; u_i^s, u_i^l 为各段中根母线电压; i 为分段区间标识; s 表示电源特性区间; l 表示负荷特性区间; P_s, P_l 分别为电源特性区间和负荷特性区间的概率信息。

该基于概率标识的广义负荷建模方法充分考虑了风电接入复杂场景下节点特性呈现不确定性的问题。通过该建模方法,不但可以获得聚类后各类别的精确广义负荷模型,还可通过测试样本泛化以检验聚类的合理性与有效性。

3.3 聚类建模流程图

通过时段序列内各最小时间长度的波动强度以及时段序列的数字特征为指标构造特征向量,采用 AP 聚类算法对时段样本数据进行聚类,然后对每个类别内样本数据,利用 RBF 神经网络按照依概率分区间的方法建立节点特性统一模型,流程图如图 3 所示。

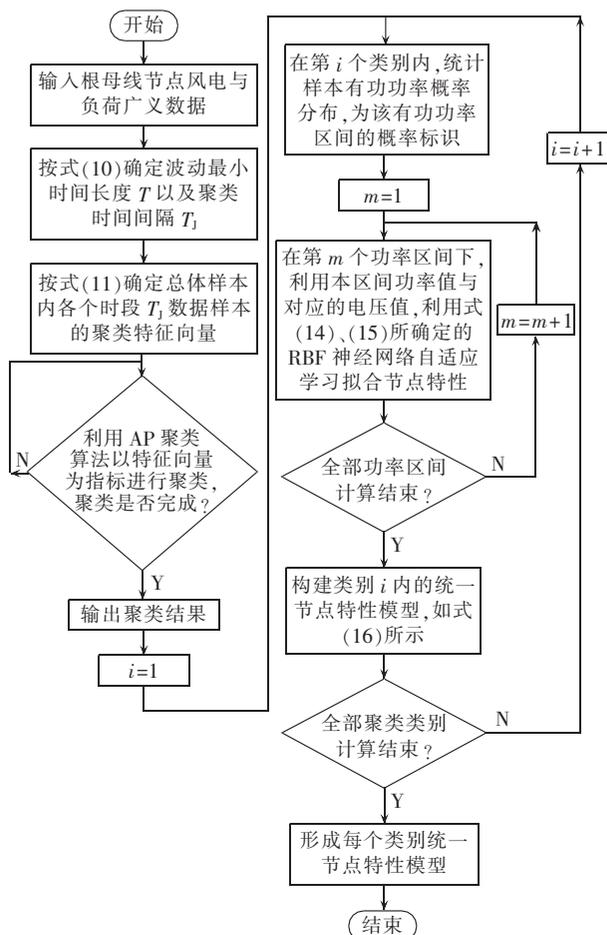


图 3 聚类建模流程图

Fig.3 Flowchart of clustering and modeling

4 算例验证及分析

本文风电数据采自 2011 年某月山东某沿海风电场实测有功运行数据,负荷数据为该地某变电站 110 kV 侧出线的功率数据。取该月前 25 天数据为训练样本,后 5 天数据为测试样本,采样间隔均为 5 min。利用风电场有功出力数据采用定功率因数的方式获得无功功率,与负荷叠加获得根母线功率,将根母线广义负荷节点视为 PQ 节点,并作为 New England 39 节点算例系统中母线 16 的功率数据,其他负荷母线功率数据满足以算例系统标准值为期望值、标准值的 5% 为标准差的正态分布,通过潮流计算获得母线 16 的电压样本,以此获得聚类与建模所需数据。

4.1 节点时段特性聚类分析

首先据式(10)确定待定时间长度内最大波动强度,结果如表 1 所示。

表 1 待定时间长度内最大波动强度
Table 1 Maximum fluctuation intensity of candidate time lengths

t/min	最大波动强度	t/min	最大波动强度
5	0	20	0.1847
10	0.2374	25	0.1538
15	0.1973		

文中合理选择波动变化率阈值 σ 为 0.2,所以满足阈值限制的最小时间长度为 15 min。

考虑聚类结果可靠性与模型的实用性因素,聚类时间间隔 T_j 取 4 h,则单日被分为 6 个连续时段。以 2011 年某月前 25 天数据为训练样本,则训练样本按时间顺序共分为 150 个时段,构造特征向量进行聚类,结果如表 2 所示。

由表 2 可以看出,训练样本数据自动划分为 4 类,聚类中心分别为第 33(类别 1)、57(类别 2)、61(类别 3)和 78(类别 4)段,其中数字为训练样本所划分时段序列数。由表 2 可知连续 2 天的时段特性也不尽相同,主观高峰低谷时段的划分方法不适用于风电接入的广义负荷场景,因此利用本文聚类方法充分统筹日时段的差异性和趋同性。

为充分说明 AP 聚类算法效果的优势,与传统 K -means 聚类算法进行对比。由于 AP 聚类算法自适应将训练样本分为 4 类,因此设定 K -means 算法的聚类数为 4。为定量分析 2 种算法的聚类效果,定义评价函数 E 如下:

$$E = \sum_{j=1}^{N_k} \sum_{i=1}^{N_c} \|X_i - m_{C_j}\|_2 \quad X_i \in C_j \quad (17)$$

其中, E 是全部样本点到所属类别聚类中心的总距离平方和; X_i 为所属类别中心 C_j 的样本点; m_{C_j} 为类别 C_j 的聚类中心; N_k 为样本类别总数; N_c 为所属类别 C_j 的样本点 X_i 数量。计算结果如表 3 所示。

表 2 聚类结果

Table 2 Results of clustering

单日号	聚类中心					
	时段 1	时段 2	时段 3	时段 4	时段 5	时段 6
1	78	78	57	57	57	57
2	78	33	33	33	33	33
3	61	33	57	57	57	78
4	78	33	33	33	33	61
5	61	61	33	33	33	33
6	61	61	33	33	33	33
7	61	61	33	33	33	33
8	61	33	57	57	57	57
9	78	78	57	57	57	78
10	33	78	57	57	57	78
11	61	33	57	57	78	78
12	33	78	78	78	78	78
13	33	61	78	78	78	78
14	33	78	57	57	57	57
15	33	61	78	57	78	78
16	78	78	57	57	57	78
17	78	78	57	57	57	78
18	33	78	57	57	57	78
19	78	33	78	57	57	57
20	78	78	78	78	57	78
21	33	61	33	33	33	78
22	33	33	78	33	33	78
23	33	33	78	78	57	57
24	78	78	57	57	57	78
25	33	61	78	78	57	57

表 3 AP 算法与 K -means 算法聚类效果对比
Table 3 Comparison of clustering effect between AP and K -means methods

类别	距离平方和	
	AP 聚类算法	K -means 聚类算法
1	6.8051	29.7961
2	5.7044	8.7638
3	1.8862	1.1391
4	13.6393	3.2695
样本总距离平方和 E	28.0350	42.9685

其中,AP 聚类算法与 K -means 聚类算法的距离平方和 E 分别为 28.0350 和 42.9685,显然采用 AP 算法聚类后各类内样本点与聚类中心距离更近,聚类效果更好。 K -means 依赖于初始聚类中心的随机选择,若初始聚类中心选择不合理,聚类结果往往较差。同时, K -means 算法对于离散和噪声数据比较敏感,少量此类数据就可产生较大影响,不适用于对负荷广义复杂场景下差别大、数据量大、较分散的样本进行聚类,而且需要聚类前人为规定聚类数目,较为主观,不能客观反映数据的类别属性。而 AP 算法不受离散和噪声数据影响,适用于大数据样本下的聚类,也无需规定聚类数,避免了主观影响,能够在数据本质属性规律的基础上,实现无监督、自适应聚类,结果更为客观。

4.2 聚类广义负荷建模分析

利用上文聚类方法得到样本数据的 4 类聚类结

果,对每类数据以概率分区间分别建立广义负荷稳态模型,为保证样本数据充足,需利用各类中全部实测样本数据,限于篇幅仅呈现类别 3 的拟合效果以证明该方法描述能力,如表 4 和图 4 所示。

表 4 类别 3 概率分布与拟合误差

Table 4 Probability distribution and fitting error of Cluster 3

分段区间	概率 P_m	拟合误差 M_m
$[-0.2, -0.1)$	0.0029	9.995×10^{-5}
$[-0.1, 0)$	0.0138	7.517×10^{-4}
$[0, 0.1)$	0.0267	5.184×10^{-4}
$[0.1, 0.2)$	0.0742	6.139×10^{-4}
$[0.2, 0.3)$	0.2342	6.256×10^{-4}
$[0.3, 0.4)$	0.3913	6.511×10^{-4}
$[0.4, 0.5)$	0.1429	7.642×10^{-4}
$[0.5, 0.6)$	0.0858	1.006×10^{-3}
$[0.6, 0.7)$	0.0283	2.649×10^{-4}

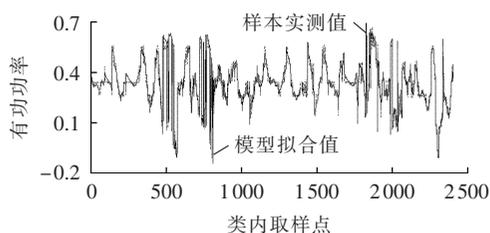


图 4 类别 3 拟合图

Fig.4 Fitting diagram of Cluster 3

由表 4 可知,本文选用的建模方法拟合误差较小,最大误差为 1.006×10^{-3} ,出现在区间 $[0.5, 0.6)$,拟合效果较好。图 4 描述了类别 3 数据的整体拟合效果(纵轴有功功率为标么值,后同),整体误差为 6.743×10^{-4} ,仅在个别点处略有偏差,整体拟合效果较好,验证了本文聚类方法的有效性。

4.3 测试样本识别验证

为验证聚类结果正确性,取测试样本中某天的全部时段进行验证,通过构建特征向量,分别与训练样本聚类生成的聚类中心的特征向量按式(18)进行欧氏距离计算,结果如表 5 所示。

$$\rho(X_j, X_c) = \sqrt{\sum_{i=1}^{N_n} [X_j(i) - X_c(i)]^2} \quad (18)$$

其中, $\rho(X_j, X_c)$ 为测试样本 X_j 与聚类中心 X_c 的欧氏距离; N_n 为向量 X_j 与 X_c 所包含的元素个数。

表 5 测试样本与各聚类中心欧氏距离

Table 5 Euclidean distance between test sample and cluster center

测试时段	欧氏距离			
	1(33)	2(57)	3(61)	4(78)
1	1.3526	1.7743	1.4626	1.2762
2	1.6154	1.8013	1.8853	1.3286
3	2.5527	1.7930	3.5957	2.2817
4	2.9662	2.1099	3.7166	2.7133
5	2.2256	2.0578	2.5578	1.8313
6	2.4777	2.4569	2.6331	1.9606

由表 5 可知,测试时段 1、2、5、6 属于类别 4,而测试时段 3、4 属于类别 2。分类结果表明,一天内各时段特性也不尽相同,这是由于不同时段内广义负荷母线风电随机波动性与负荷时变性叠加相互作用而使广义负荷所表现出的功率特性不同,从而产生了负荷母线特性的时段差异性。为充分验证聚类方法的有效性,分别用 4 类模型对测试样本时段 3 进行拟合,拟合效果如图 5 所示。

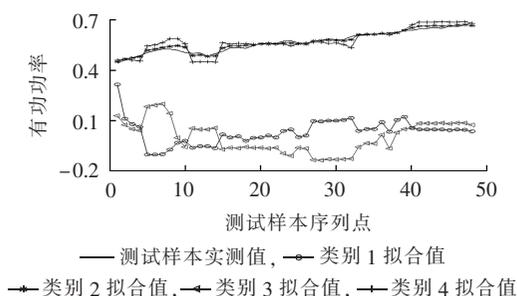


图 5 待检验时段拟合效果

Fig.5 Fitting effects of different models

可知,4 类模型拟合误差分别为 0.2967 、 4.3×10^{-4} 、 0.3342 、 9.86×10^{-4} 。显然,利用类别 2 拟合效果较好,类别 4 拟合效果次之,类别 1、3 拟合脱离测试样本实测值,误差大,通过特性综合再次证明了本文提出聚类方法的正确性。用本文方法与传统最小二乘法^[20]对训练样本中某天数据进行建模,拟合效果如图 6 所示。

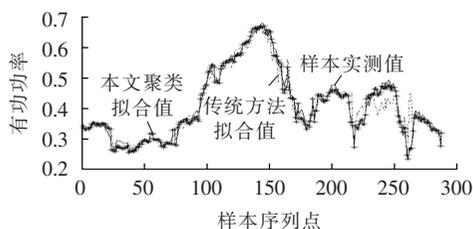


图 6 本文方法与传统建模方法结果比较

Fig.6 Comparison between proposed and traditional modeling methods

可见,当广义负荷节点功率特性变化波动剧烈时,传统方法难以准确拟合,而本文方法先对数据进行聚类处理,利用同类数据建立广义负荷模型,建模结果更加精确。

5 基于 AP 聚类算法的广义负荷稳态建模在风险分析中的应用

针对日时段样本经过本文所提出聚类方法聚类后得到的广义负荷聚类类别信息,引入带有概率信息的 RBF 神经网络模型进行建模,能够在反映节点时段特性的基础上对不确定问题进行全局描述。因此,从风险评估角度综合考虑支路潮流临近度、节点电压临近度和场景发生概率,分析系统在稳态运行情

况下的潜在风险。

本文以 IEEE 5 节点系统为例,将前文建立的 4 类广义负荷概率模型分别作为母线 3 的节点特性模型,修正功率不平衡量如下:

$$\Delta P_i = \sum_{k=1}^N w_{pk} R_{pk}(u, \rho) - U_i \sum_{j \in i} U_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}) \quad (19)$$

$$\Delta Q_i = \sum_{k=1}^N w_{qk} R_{qk}(u, \rho) - U_i \sum_{j \in i} U_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij}) \quad (20)$$

其中, N 为区间样本数; u 为节点电压输入量; w_{pk} 和 $R_{pk}(u)$ 分别为求解有功修正量时第 p 个输出层连接权值和第 p 个隐含层径向基函数; w_{qk} 和 $R_{qk}(u)$ 分别为求解无功修正量时第 q 个输出层连接权值和第 q 个隐含层径向基函数; U_i 和 U_j 分别为节点 i 和节点 j 处的电压值; G_{ij} 、 B_{ij} 和 θ_{ij} 分别为线路电导、电纳和节点电压相角差。

按式(19)和(20)分别修正雅可比矩阵中对角元素,分别进行类别内各分段功率场景下的潮流计算,获取支路潮流和节点电压,定义支路潮流失稳临近度 P_c 和节点电压失稳临近度指标 U_c 考核系统运行的潜在风险。根据风险评估定义,风险度指标 δ_p 、 δ_U 为发生可能性与严重度乘积,如式(21)所示。

$$\begin{cases} P_c = \left| \frac{P_{mm} - (P_{mm}^{\max} + P_{mm}^{\min})/2}{(P_{mm}^{\max} - P_{mm}^{\min})/2} \right| \\ U_c = \left| \frac{U - (U^{\max} + U^{\min})/2}{(U^{\max} - U^{\min})/2} \right| \\ \delta_p = P_m P_c \\ \delta_U = P_m U_c \end{cases} \quad (21)$$

其中, P_{mm} 为支路功率; P_{mm}^{\max} 、 P_{mm}^{\min} 分别为支路功率上、下限,分别为 2.5、-2.5 p.u.; U 为节点电压; U^{\max} 、 U^{\min} 分别为节点电压上、下限,分别取 1.1、0.9 p.u.; P_m 为类内功率区间 m 出力概率,即发生的可能性指标。

根据式(21),限于篇幅以类别 4 为例,风险分析结果如表 6 所示。

表 6 节点 3 电压风险评估

Table 6 Voltage risk assessment for node 3

功率区间	概率	节点电压	邻近度	风险度
[-0.2, -0.1)	0.0029	0.9774	0.2258	6.5857×10^{-4}
[-0.1, 0)	0.0138	0.9667	0.3329	0.0046
[0, 0.1)	0.0267	0.9354	0.6458	0.0172
[0.1, 0.2)	0.0742	1.0202	0.2021	0.0150
[0.2, 0.3)	0.2342	0.9968	0.0324	0.0076
[0.3, 0.4)	0.3913	0.9997	0.0029	0.0011
[0.4, 0.5)	0.1429	0.9908	0.0924	0.0132
[0.5, 0.6)	0.0858	0.9899	0.1011	0.0087
[0.6, 0.7)	0.0283	0.9882	0.1177	0.0033

表 6 从节点电压的角度,综合考虑 U_c 与功率区间发生可能性 P_m ,得出其风险度 δ_U ,其中 [0, 0.1)、[0.1, 0.2) 的 δ_U 分别为 0.0172、0.015; 表 7 从支路潮流角度,综合考虑 P_c 与功率区间发生可能性 P_m ,得

表 7 母线 2、3 间支路潮流风险评估

Table 7 Power flow risk assessment for branch between bus 2 and bus 3

功率区间	概率	支路潮流	邻近度	风险度
[-0.2, -0.1)	0.0029	0.7757	0.3103	9.0494×10^{-4}
[-0.1, 0)	0.0138	0.7276	0.2910	0.0040
[0, 0.1)	0.0267	0.4547	0.1819	0.0049
[0.1, 0.2)	0.0742	0.6496	0.2598	0.0193
[0.2, 0.3)	0.2342	0.6070	0.2428	0.0569
[0.3, 0.4)	0.3913	0.5692	0.2277	0.0891
[0.4, 0.5)	0.1429	0.5292	0.2117	0.0303
[0.5, 0.6)	0.0858	0.4815	0.1926	0.0165
[0.6, 0.7)	0.0283	0.4465	0.1786	0.0051

出其风险度 δ_p ,其中 [0.2, 0.3)、[0.3, 0.4) 的 δ_p 分别为 0.0569、0.0891。因此,当出现上述功率区间时应通过调度控制措施避免系统越限而造成损失。

对于文中广义负荷聚类形成的 4 个类别,根据式(22),综合考虑每个类别内的每个功率区间对节点电压的影响进行风险分析统计。

$$\delta_i^j = \sum_{m=1}^N P_{i,m} U_c^{i,m} \quad (22)$$

其中, $P_{i,m}$ 为类别 i 第 m 个功率区间概率值; $U_c^{i,m}$ 为类别 i 内第 m 个功率区间电压失稳邻近度; N_i 为类别 i 的功率区间分段数; δ_i^j 为类别 i 的整体电压风险度。

对 4 类聚类类别中节点 3 风险分析结果进行对比,如表 8 所示。

表 8 风险分析对比

Table 8 Comparison of voltage risk

类别	电压风险	类别	电压风险
1	0.1510	3	0.1103
2	0.1852	4	0.0714

由表 8 分析可知,类别 2 总体电压风险较高,这是由于类别 2 内风电接入水平较高,系统不确定性因素较强,节点越限风险大。因此,当判别时段特性样本归属于类别 2 时,应特别注意,否则易引起系统运行偏差,从而带来经济、安全隐患。

6 结论

本文将客观的、适用于复杂数据场景的 AP 算法应用于电力系统广义负荷特性聚类,通过实测样本空间的直观聚类结果与特性综合,并与 K-means 聚类算法比较,说明了该算法的有效性和优越性。

a. 提出了按时间段划分并以时段内波动强度序列以及统计量构造特征向量的聚类方法。该聚类方法能够反映样本数据的日时段特性规律,不仅可以充分反映日时段内的差异性和趋同性,而且可以直观表达日间节点特性的差异性。

b. 通过模式匹配判断样本所属类别,利用广义负荷建模检验聚类有效性。仿真结果表明,测试样本采用所属类别的模型拟合效果较好,因此通过本

文 AP 聚类后综合能够得到精确广义负荷模型。

c. 将聚类后综合得到精确广义负荷模型应用于风电接入后的风险仿真计算分析,结果指出了高风险节点功率区间,可为系统决策提供参考。

参考文献:

- [1] 李欣然,惠金花,钱军,等. 风力发电对配网侧负荷建模的影响[J]. 电力系统自动化,2009,33(13):89-94.
LI Xinran,HUI Jinhua,QIAN Jun,et al. Impact of wind power generation on load modeling in distribution network[J]. Automation of Electric Power Systems,2009,33(13):89-94.
- [2] 孙东磊,韩学山,李文博. 风储共存于配网的动态优化潮流分析[J]. 电力自动化设备,2015,35(8):110-117.
SUN Donglei,HAN Xueshan,LI Wenbo. Analysis of dynamic optimal power flow for distribution network with wind power and energy storage[J]. Electric Power Automation Equipment,2015,35(8):110-117.
- [3] ENGELHARDT S,ERLICH I,FELTES C,et al. Reactive power capability of wind turbines based on doubly fed induction generators[J]. IEEE Trans on Energy Conversion,2011,26(1):364-372.
- [4] 黄学良,刘志仁,祝瑞金,等. 大容量变速恒频风电机组接入对电网运行的影响分析[J]. 电工技术学报,2010,25(4):142-149.
HUANG Xueliang,LIU Zhiren,ZHU Ruijin,et al. Impact of power system integrated with large capacity of variable speed constant frequency wind turbines[J]. Transactions of China Electrotechnical Society,2010,25(4):142-149.
- [5] 杨悦,李国庆,王振浩. 基于可信性理论的含风电电力系统电压稳定概率评估[J]. 电力自动化设备,2014,34(12):6-12.
YANG Yue,LI Guoqing,WANG Zhenhao. Probabilistic voltage stability assessment based on credibility theory for power system with wind farm[J]. Electric Power Automation Equipment,2014,34(12):6-12.
- [6] 章健. 电力系统负荷建模方法的研究[D]. 北京:华北电力大学,1997.
ZHANG Jian. Studies on modeling methodology of electric loads [D]. Beijing:North China Electric Power University,1997.
- [7] 张红斌,贺仁睦,刘应梅. 基于 KOHONEN 神经网络的电力系统负荷特性聚类与综合[J]. 中国电机工程学报,2003,23(5):1-5.
ZHANG Hongbin,HE Renmu,LIU Yingmei. The characteristics clustering and synthesis of electric dynamic loads based on KOHONEN neural network[J]. Proceedings of the CSEE,2003,23(5):1-5.
- [8] 白雪峰,蒋国栋. 基于改进 K-means 聚类算法的负荷建模及应用[J]. 电力自动化设备,2010,30(7):80-83.
BAI Xuefeng,JIANG Guodong. Load modeling based on improved K-means clustering algorithm and its application[J]. Electric Power Automation Equipment,2010,30(7):80-83.
- [9] 李培强,李欣然,陈凤,等. 模糊聚类在统计综合合法负荷建模中的应用[J]. 电力自动化设备,2003,23(5):43-45.
LI Peiqiang,LI Xinran,CHEN Feng,et al. Application of fuzzy clustering in component-based modeling approach [J]. Electric Power Automation Equipment,2003,23(5):43-45.
- [10] 鞠平,王耀,项丽,等. 考虑特殊负荷的宁夏电网负荷建模[J]. 电力自动化设备,2012,32(8):1-4.
JU Ping,WANG Yao,XIANG Li,et al. Load modeling for Ningxia grid special loads[J]. Electric Power Automation Equipment,2012,32(8):1-4.
- [11] 鞠平,金艳,吴峰,等. 综合负荷特性的分类综合方法及其應用[J]. 电力系统自动化,2004,28(1):64-68.
JU Ping,JIN Yan,WU Feng,et al. Studies on classification and synthesis of composite dynamic loads[J]. Automation of Electric Power Systems,2004,28(1):64-68.
- [12] 黄梅,贺仁睦,杨少兵. 模糊聚类在负荷实测建模中的应用[J]. 电网技术,2006,30(14):49-52.
HUANG Mei,HE Renmu,YANG Shaobing. Application of fuzzy clustering in measurement-based load modeling[J]. Power System Technology,2006,30(14):49-52.
- [13] 李培强,李欣然,陈辉华,等. 基于减法聚类的模糊神经网络负荷建模[J]. 电工技术学报,2006,21(9):2-6.
LI Peiqiang,LI Xinran,CHEN Huihua,et al. Fuzzy neural network load modeling based on subtractive clustering[J]. Transactions of China Electrotechnical Society,2006,21(9):2-6.
- [14] 张旭,梁军,负志皓,等. 考虑风电接入不确定性的广义负荷建模[J]. 电力系统自动化,2014,38(20):61-67.
ZHANG Xu,LIANG Jun,YUN Zhihao,et al. Generalized load modeling and application considering uncertainty of wind power integration[J]. Automation of Electric Power Systems,2014,38(20):61-67.
- [15] WANG Z,BIAN S,LIU Y,et al. The load characteristics classification and synthesis of substations in large area power grid[J]. International Journey of Electrical Power & Energy Systems,2013(48):71-82.
- [16] ORDONEZ C,OMIECINSKI E. Efficient disk-based K-means clustering for relational databases[J]. IEEE Transactions on Knowledge and Data Engineering,2004,16(8):909-921.
- [17] FREY B J,DUECK D. Clustering by passing messages between data points[J]. Science,2007,315(5814):972-976.
- [18] 张学清,梁军,董晓明. 基于累积指数的电网电压稳定性能的评估[J]. 电工技术学报,2012,27(7):235-241.
ZHANG Xueqing,LIANG Jun,DONG Xiaoming. An assessment for voltage stability performance of power grid based on cumulating index[J]. Transactions of China Electrotechnical Society,2012,27(7):235-241.
- [19] 翟学锋,卫志农,范立新,等. 基于相关向量机的发电机进相能力建模[J]. 电力自动化设备,2015,35(3):146-151.
ZHAI Xuefeng,WEI Zhinong,FAN Lixin,et al. Generator leading phase capability model based on relevance vector machine[J]. Electric Power Automation Equipment,2015,35(3):146-151.
- [20] 徐兵. 基于在线数据的负荷建模研究[D]. 济南:山东大学,2013.
XU Bing. Research on load modeling based on on-line data [D]. Ji'nan:Shandong University,2013.

作者简介:



褚壮壮

褚壮壮(1989—),男,山东济宁人,博士研究生,主要研究方向为电力系统运行与控制(E-mail:chulsun@163.com);

梁军(1956—),男,山东济南人,教授,博士研究生导师,主要研究方向为电力系统运行与控制;

张旭(1989—),女,山东滨州人,硕士研究生,主要研究方向为电力系统运行与控制;

董晓明(1980—),男,山东济南人,副研究员,主要研究方向为电力系统运行与控制。

Modeling of generalized load steady-state characteristics based on affinity propagation clustering algorithm and its application

CHU Zhuangzhuang¹, LIANG Jun¹, ZHANG Xu², DONG Xiaoming¹, ZHANG Yongliang³

(1. Key Laboratory of Power System Intelligent Dispatch and Control, Ministry of Education, Shandong University, Ji'nan 250061, China; 2. State Grid Jinan Power Company, Ji'nan 250012, China; 3. State Grid Binzhou Power Company, Binzhou 256600, China)

Abstract: Since the grid-connection of large-scale renewable energy increases the uncertainty of nodal power flow, the source and load characteristics of node become unclear, which brings new challenges to the generalized load modeling. A method based on AP (Affinity Propagation) clustering algorithm is proposed for the generalized load steady-state characteristics clustering. The characteristic analysis is carried out for the root bus power data of generalized load node and the minimum time period of power fluctuation sequence is determined based on the fluctuation intensity theory of dynamics. The daily periodic feature vector is formed by the fluctuation intensity sequence of sample with the minimum time length in each period and the digital feature index of sample data within the period, with which as the clustering index, the AP clustering algorithm is applied to adaptively adjust the cluster number and center. The probabilistic information is introduced to the generalized load modeling method, which is then used to build the model for different clusters and verify the accuracy of clustering algorithm. Simulative analysis shows that, the clustering algorithm adaptively determines the clusters to fully reflect the daily periodic features, and an accurate generalized load model is obtained by the characteristic synthesis, which is applied to the simulative analysis of wind-power integration risk.

Key words: load modeling; load characteristic; affinity propagation clustering algorithm; fluctuation intensity; time period characteristics; generalized load characteristic; clustering; RBF neural network

(上接第 114 页 continued from page 114)

Time distribution of weather-related transmission line failure and its fitting

WANG Jian¹, XIONG Xiaofu¹, LI Zhe², LIANG Yun², WENG Shijie¹

(1. State Key Laboratory of Power Transmission Equipment & System Security and New Technology, Chongqing University, Chongqing 400044, China; 2. Electric Power Research Institute of State Grid Henan Electric Power Company, Zhengzhou 450052, China)

Abstract: Neither accurate time distribution of transmission line failure rate nor accurate probability distribution description of forced outage time exists in the risk analysis and reliability assessment of power grid. A method for calculating the time-varying failure rate based on the historical data of monthly failure rate for several years is proposed. The time distribution of monthly failure rate is fitted with Fourier, Gaussian and Weibull functions based on the calculated monthly failure rates of same period to get the monthly failure rate function for every month, which is used to simulate the time-dependent failure properties of transmission line for predicting the failure rate of a certain time period in the future. In like manner, the probability distribution of weather-related forced outage time is fitted and modelled for transmission line. Case analysis verifies the effectiveness of the proposed method.

Key words: transmission line; meteorological disaster; time-varying failure rate; forced outage time; time distribution