

基于分裂层次半监督谱聚类算法的风电场机群划分方法

林 俐, 潘险险

(华北电力大学 新能源电力系统国家重点实验室, 北京 102206)

摘要: 在谱聚类技术的基础上, 面向风电场动态等值建模, 提出一种基于分裂层次半监督谱聚类算法的风电场机群划分方法。首先根据风电机组的实测运行数据, 构造一个可以体现原始数据结构且能为分类提供更多有效信息的特征向量矩阵 Y 。进而利用获取的部分样本组的先验信息, 采用自顶向下的簇分裂策略, 对 Y 中的样本组进行半监督聚类划分, 得到风电机组的机群划分结果, 采用容量加权法计算各机群等值风电机组的参数, 建立风电场的动态等值模型。以某实际风电场为例进行仿真验证, 结果表明, 采用所提方法建立的动态等值模型与详细模型较接近, 能够较准确地反映风电场的动态响应特性。

关键词: 分裂层次; 半监督; 谱聚类; 风电场; 机群划分; 聚类算法; 动态等值

中图分类号: TM 614

文献标识码: A

DOI: 10.16081/j.issn.1006-6047.2015.02.002

0 引言

随着能源和环境问题的日益严重, 风力发电越来越受到世界各国的重视。由于风能具有随机性、间歇性和不稳定性的特点, 随着风电机组单机容量和风电场规模的不断扩大, 风电并网对电力系统稳定性的影响愈发显著^[1-3]。为准确分析和评价大容量风电场与电力系统之间的相互作用和影响, 研究并寻求合适的风电场动态等值方法对含风电场的电力系统动态稳定仿真分析具有重要意义。

国内外学者对风电场动态等值模型进行了许多研究, 通常假设所有风电机组的输入风速相同, 将整个风电场等值为一台风电机组。然而对于地形复杂、机组排列不规则的大型风电场, 受风速的尾流效应和时滞等因素的影响, 风电机组的运行状态并不相同, 个别机组特性差异很大^[4-7]。文献[7]通过研究认为尾流效应会造成风电场的风能损失, 时滞会平滑风电场输出功率, 使并网点功率变化率变小, 尾流效应和时滞将对风电场输出特性产生重要影响, 使用单机表征法通常会产生较大误差。为提高风电场等值模型的精度, 近年来, 一些学者借鉴传统电力系统动态等值中的同调等值法, 基于实测运行数据并采用某些聚类算法, 以机组具有相同或相近运行点为划分原则进行机群划分, 用多机表征法对风电场进行等值, 实测运行数据一定受到地形、机组排列、尾流效应和迟滞等因素的影响, 也是各风电机组内部复杂的物理过程的结果^[8]。就聚类算法而言, K-means 聚类算法是最传统、经典的聚类算法, 文献[9-13]基于 K-means 聚类

算法建立了风电场动态等值的单机表征模型。然而, K-means 算法只能识别凸球形分布的数据, 当样本空间不为凸时, 算法可能会陷入局部最优^[14]。谱聚类算法能有效解决这一问题, 它建立在谱图理论基础上, 能在任意形状的样本空间上聚类且收敛于全局最优解。文献[15]提出基于扩散映射理论的谱聚类算法对风电场内的机组进行机群划分, 但是该方法是一种无监督的学习算法, 仅利用样本层面的信息来进行聚类, 当所定义的聚类目标函数不适合数据本身时, 该方法不能达到较好的机群划分效果。

据此, 本文在项目组前期研究^[15]基础上, 面向风电场动态等值建模, 提出一种基于分裂层次半监督谱聚类算法的风电场机群划分方法。首先基于风电场内所有风电机组的实测运行数据, 根据欧氏距离构造相似性矩阵, 进而构建拉普拉斯矩阵。然后对拉普拉斯矩阵进行谱特征分析, 选择合适的特征值, 构造一个可以体现原始数据空间结构且能为分类提供更多有效信息的特征向量空间。在对该空间中的样本组进行聚类划分时, 采用自顶向下的簇分裂策略, 从数据内部固有的分布特性, 利用获取的部分样本组的先验信息, 进行半监督聚类划分, 进而得到风电机组的机群划分结果。最后采用算例仿真, 通过将本文基于分裂层次半监督谱聚类算法建立的多机表征模型与传统的单机等值模型、基于无监督谱聚类算法建立的等值模型与详细模型对比, 验证采用本文提出的机群划分方法建立的风电场动态模型的有效性。

1 基于分裂层次半监督谱聚类算法的风电机组机群划分

作为一种有效的数据分析方法, 谱聚类算法已被广泛应用于各个领域, 相对于传统的聚类算法而言, 谱聚类算法能在任意形状的样本空间上聚类且收敛

收稿日期: 2014-05-04; 修回日期: 2014-12-15

基金项目: 国家自然科学基金重大项目(51190103); 国家科技支撑计划项目(2013BAA02B01)

Project supported by the Key Program of National Natural Science Foundation of China(51190103) and the National Key Technology R & D Program of China(2013BAA02B01)

于全局最优解^[16]。谱聚类算法首先根据数据样本组之间的欧氏距离构造相似性矩阵,进而构造拉普拉斯矩阵,通过对拉普拉斯矩阵进行谱特征分析,构造出归一化的特征向量空间,该数据空间可以体现原始数据空间的结构,且使原数据空间中样本组之间的相似性关系表现得更加直观、明显。通过对该特征向量空间进行聚类划分,即可得到原数据空间中样本组的划分结果。

然而,在对特征向量空间进行聚类划分时,很容易忽视样本组内部固有的分布特性,若能利用部分样本组的先验信息来指导聚类划分过程,即进行半监督聚类划分,可以有效地改善聚类效果^[17]。

1.1 构建归一化的特征向量空间

假设风电场中共有 $n+m$ 台风电机组,若某个时段内有 n 台风电机组并网运行, m 台风电机组由于某些原因而与电网脱离,则将 m 台离网的风电机组数据剔除,选取风电场在该时段内并网运行的风电机组的实测数据作为样本,在该时段内实测数据的采样点数为 t 个,将 n 台风电机组的实测数据建立样本矩阵:

$$\mathbf{V}=(\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_n)^T=(v_{i,j})_{n \times t}=\begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,t} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,t} \\ \vdots & \vdots & & \vdots \\ v_{n,1} & v_{n,2} & \cdots & v_{n,t} \end{bmatrix} \quad (1)$$

其中, $v_{i,j}$ 表示第 i 台风电机组在第 j 个时刻的实测运行数据。把向量 $\mathbf{V}_i$ 看作空间 $\mathbf{V}$ 的一个样本组,则 $\mathbf{V}$ 中样本组数量为 n ,维数为 t 。由实测数据样本矩阵 $\mathbf{V}$ 构建 n 台风电机组两两之间的欧氏距离矩阵 $\mathbf{H}=(H_{u,w})_{n \times n \odot}$

$$H_{u,w}=d(\mathbf{V}_u, \mathbf{V}_w)=\sqrt{\sum_{j=1}^t (v_{u,j}-v_{w,j})^2} \quad (2)$$

其中, $u, w=1, 2, \cdots, n$; $d(\mathbf{V}_u, \mathbf{V}_w)$ 表示第 u 台风电机组与第 w 台风电机组实测数据样本组之间的欧氏距离; $\mathbf{H}$ 是主对角元素为 0 的对称阵。

这里采用高斯函数构建 $\mathbf{H}$ 的相似性矩阵,考虑到谱聚类算法对高斯函数尺度参数的取值非常敏感^[18],为消除尺度参数对聚类效果的影响,本文采用自适应尺度参数代替固定的尺度参数,构造相似性矩阵 $\mathbf{A}$:

$$\mathbf{A}=(A_{u,w})_{n \times n}=\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,n} \end{bmatrix} \quad (3)$$

$$A_{u,w}=\begin{cases} 0 & u=w \\ e^{-H_{u,w}^2/(\sigma_u \sigma_w)} & u \neq w \end{cases} \quad (4)$$

$$\sigma_u=\frac{1}{r} \sum_{l=1}^r d(\mathbf{V}_u, \mathbf{V}_l) \quad (5)$$

其中, σ_u 和 σ_w 为自适应尺度参数, σ_u 为与 $\mathbf{V}_u$ 欧氏距

离最小的 r 个样本组的平均欧氏距离,通常 r 取 3~5; $\mathbf{V}_l$ 为与 $\mathbf{V}_u$ 欧氏距离较小的第 l 个样本组; σ_w 类似。 σ_u 和 σ_w 主要用来控制 $\mathbf{V}_u$ 和 $\mathbf{V}_w$ 之间的欧氏距离对相似性矩阵 $\mathbf{A}$ 中的元素 $A_{u,w}$ 的影响^[19]。

基于相似性矩阵 $\mathbf{A}$ 建立度矩阵 $\mathbf{D}=(D_{u,w})_{n \times n \odot}$

$$D_{u,w}=\begin{cases} \sum_{w=1}^n A_{u,w} & u=w \\ 0 & u \neq w \end{cases} \quad (6)$$

进而可以构建拉普拉斯矩阵 $\mathbf{L}$:

$$\mathbf{L}=\mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} \quad (7)$$

谱聚类算法中,拉普拉斯矩阵 $\mathbf{L}$ 是一个非常重要的矩阵,该矩阵的所有特征值连同其重数构成的集合为矩阵 $\mathbf{L}$ 的谱,其最大特征值为矩阵 $\mathbf{L}$ 的谱半径。将矩阵 $\mathbf{L}$ 的特征值从大到小排列, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$,可以证明对于 k 个理想的彼此分离簇的有限数据集,矩阵 $\mathbf{L}$ 的前 k 个最大特征值为 1,第 $k+1$ 个特征值则严格小于 1,二者之间的差距取决于这 k 个聚类的分布情况^[20]。但是对于普通分布的样本数据,其特征值未必会严格符合这一关系^[21]。因此,需要采用合适的方法来找到前 k 个主导特征值。

定义相邻特征值的差值为本征间隙 δ_e :

$$\delta_e=\lambda_e-\lambda_{e+1} \quad e=1, 2, \cdots, n-1 \quad (8)$$

根据矩阵摄动理论,本征间隙越大,选取的 k 个特征向量所构成的子空间就越稳定^[22],当样本组簇内分布越紧密、簇间分布越分离时,相应的最大本征间隙值越大。因此,可以根据本征间隙找出前 k 个主导特征值,确定聚簇的个数 k 。

若 $\delta_e(e=1, 2, \cdots, n-1)$ 最大,则取聚簇的个数 $k=e$,前 k 个主导特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 。设 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 所对应的特征向量为 $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_k$ (重复特征值选择正交的特征向量),则特征向量矩阵 $\mathbf{X}$ 为:

$$\mathbf{X}=(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_k)=(x_{f,g})_{n \times k}=\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{bmatrix} \quad (9)$$

对 $\mathbf{X}$ 进行归一化,记归一化后的矩阵为 $\mathbf{Y}$:

$$y_{f,g}=\frac{x_{f,g}}{\sqrt{\sum_{g=1}^k x_{f,g}^2}} \quad (10)$$

矩阵 $\mathbf{Y}$ 可以体现原始数据空间 $\mathbf{V}$ 的结构,且能为分类提供更多有效的信息,使原数据样本组之间的相似性关系表示得更加直观、明显,通过对 $\mathbf{Y}$ 中样本组进行聚类划分,划分结果组后映射回原数据空间。

1.2 基于分裂层次半监督的聚类划分

在半监督聚类中,监督信息分为 2 种:一种是由用户提供的数据对象的类属信息,获得这部分信息比较困难;另一种是聚类数据对象之间的一些内在关系,这部分信息较容易获取,因此,本文利用获取的这部分先验信息作为监督信息进行半监督聚类。

把 Y 的每一行 Y_i 看作空间 $R^{n \times k}$ 中的一个样本组 (样本组数量为 n , 样本维数为 k), 在对 Y 中的样本组进行聚类划分时, 采用自顶向下的簇分裂策略, 将部分样本组的先验信息形成半监督信息, 指导其后的聚类划分过程, 使各个簇内的样本组相似度更大, 簇间的样本组相似度更小。

计算 Y 中任意 2 个样本组之间的欧氏距离:

$$d(Y_p, Y_q) = \sqrt{\sum_{g=1}^k (y_{p,g} - y_{q,g})^2} \quad p, q = 1, 2, \dots, n \quad (11)$$

当 $d(Y_p, Y_q)$ 取得最大值时, 令所对应的样本组 $Y_p = B_1, Y_q = B_2$, 此时, B_1 和 B_2 构成一对数据对象对, 形成半监督信息, 指导其后的聚类划分过程。将 B_1 和 B_2 当作聚类中心, 把 Y 中剩余的 $n-2$ 个样本组根据欧氏距离划分为 2 个簇, 记以 B_1 和 B_2 为聚类中心的簇分别为 c_1^1 和 c_2^1 (c_j^i 表示第 i 次划分时以 B_j 为聚类中心的簇), 具体的划分方法如下。

a. 对 Y 中剩余的 $n-2$ 个样本组中任意一个样本组 Y_i , 若 $d(Y_i, B_1) < d(Y_i, B_2)$, 则将 Y_i 划分进 c_1^1 中, 否则将 Y_i 划分进 c_2^1 中。遍历完所有样本组, 将所有的样本组划分为 2 个簇, 完成第 1 次划分。

b. 寻找第 3 个聚类中心, 记:

$$z_1^1 = \max \{d(Y_i, B_1), Y_i \in c_1^1\} \quad (12)$$

其中, z_j^i 表示簇 c_j^i 内所有样本组与该簇聚类中心最大的欧氏距离。当 $d(Y_i, B_1) (Y_i \in c_1^1)$ 取得最大值时, 令所对应的样本组 $Y_i = G_{3,10}$ 。

记:

$$z_2^1 = \max \{d(Y_i, B_2), Y_i \in c_2^1\} \quad (13)$$

当 $d(Y_i, B_2) (Y_i \in c_2^1)$ 取得最大值时, 令所对应的样本组 $Y_i = G_{3,2}$, 若 z_1^1 较大, 取 $B_3 = G_{3,1}$ (B_3 与 B_1 构成一对数据对象对, 形成半监督信息), 否则, 取 $B_3 = G_{3,2}$ (B_3 与 B_2 构成一对数据对象对, 形成半监督信息)。

取 B_3 为第 3 个聚类中心, 将 Y 中剩余的 $n-3$ 个样本组根据欧氏距离划分为 3 个簇, 记当前以 B_1, B_2, B_3 为聚类中心的簇分别为 c_1^2, c_2^2, c_3^2 , 完成第 2 次聚类划分。

c. 寻找第 4 个聚类中心, 记:

$$z_3^2 = \max \{d(Y_i, B_1), Y_i \in c_1^2\} \quad (14)$$

当 $d(Y_i, B_1) (Y_i \in c_1^2)$ 取得最大值时, 令所对应的样本组 $Y_i = G_{4,10}$ 。

记:

$$z_4^2 = \max \{d(Y_i, B_2), Y_i \in c_2^2\} \quad (15)$$

当 $d(Y_i, B_2) (Y_i \in c_2^2)$ 取得最大值时, 令所对应的样本组 $Y_i = G_{4,20}$ 。

记:

$$z_5^2 = \max \{d(Y_i, B_3), Y_i \in c_3^2\} \quad (16)$$

当 $d(Y_i, B_3) (Y_i \in c_3^2)$ 取得最大值时, 令所对应的样本组 $Y_i = G_{4,3}$, 记 $z_h^2 = \max \{z_i^2, i = 1, 2, 3\}$, 若 z_h^2 最大, 取

$B_4 = G_{4,h}$, 此时 B_4 与 B_h 构成一对数据对象对, 形成半监督信息, 指导其后的聚类划分过程。

选 B_4 为第 4 个聚类中心, 将 Y 中剩余的 $n-4$ 个样本组根据欧氏距离划分为 4 个簇, 记当前以 B_1, B_2, B_3, B_4 为聚类中心的簇分别为 $c_1^3, c_2^3, c_3^3, c_4^3$, 完成第 3 次聚类划分。

d. 依此类推, 直到将 Y 中所有的样本组划分为 k 个簇, 完成整个聚类划分过程, 当样本组 Y_i 被划分进第 s 个簇 $c_s^k (s = 1, 2, \dots, k)$, 则表示样本组 Y_i 被划分到第 s 个簇 $c_s^k (s = 1, 2, \dots, k)$, 即第 i 台风电机组被划分到了第 $s (s = 1, 2, \dots, k)$ 个机群。

2 算例仿真分析

2.1 某实际风电场的机群划分

根据上述风电场机群划分方法, 编写了 MATLAB 聚类程序。算例为某实际风电场, 该风电场内共有 33 台额定容量为 1.5 MW 的双馈风电机组, 风电场总装机容量为 49.5 MW, 该风电场位于坝上丘陵地区, 所处位置有连绵不断的低矮山丘, 为了适合地势变化, 风电机组位置平面图如图 1 所示。

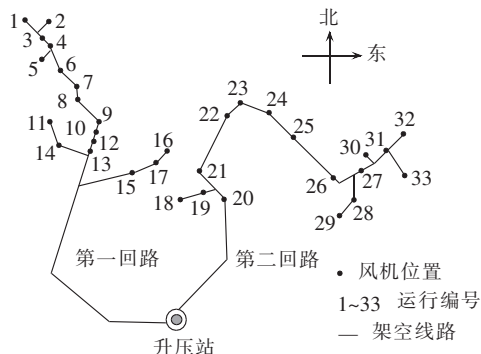


图 1 某实际风电场机组位置平面图

Fig.1 Wind turbines layout of an actual wind farm

风速是表征风电机组运行特性的输入特征量, 反映风在机组上的作用情况, 而各风电机组的输出功率是其风速经过复杂的风能转化为电能这一物理过程的最终结果, 是各风电机组所受风速、所在地理位置的地形以及机组实际运行性能等影响因素的综合反馈, 故本文选取风电机组实测功率数据为特征量进行机群划分。这里选取 2010 年 6 月 1 日至 2010 年 6 月 30 日风电场内各风电机组的 10 min 实测功率数据进行分析, 当某风电机组停运时, 为减少坏数据的影响, 将该时段内所有机组的实测运行数据剔除, 采用本文提出的方法进行机群划分, 根据实测有功功率的划分结果如表 1 所示。

由于本文采用的是每 10 min 实测运行数据, 一个月的实测数据时间点过多, 为显示方便, 截取其中 5 h 的实测有功功率数据来说明结果, 图 2—5 分别是 4 个机群在这 5 h 内的实测有功功率数据分布图。

表 1 基于实测有功功率数据的机群划分结果
Table 1 Result of wind turbine grouping based on measured active power

机群号	机群内包含的风机编号
1	1、15、21、25、28、30、32、33
2	3、5、7、11、29、31
3	2、6、9、10、13、16、17、18、19、20、22、23、24、26、27
4	4、8、12、14

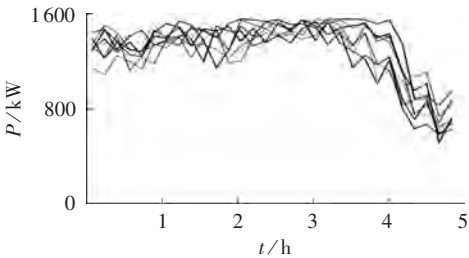


图 2 机群 1 内机组实测有功功率数据分布图
Fig.2 Line chart of measured active power for wind turbines of group 1

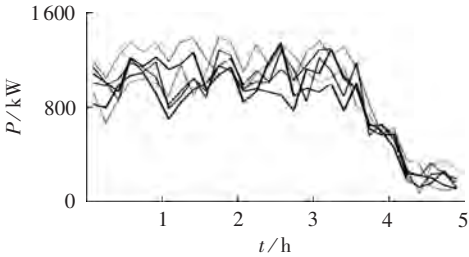


图 3 机群 2 内机组实测有功功率数据分布图
Fig.3 Line chart of measured active power for wind turbines of group 2

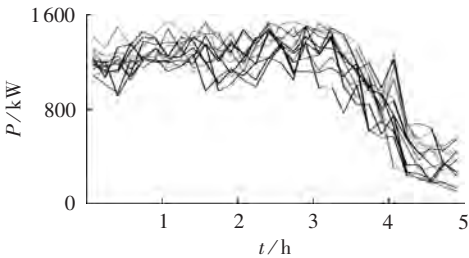


图 4 机群 3 内机组实测有功功率数据分布图
Fig.4 Line chart of measured active power for wind turbines of group 3

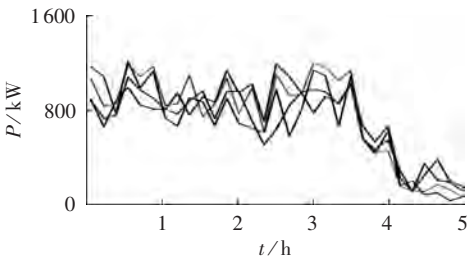


图 5 机群 4 内机组实测有功功率数据分布图
Fig.5 Line chart of measured active power for wind turbines of group 4

从图 2—5 可以看出,同群内的有功功率数据分布较为接近,与机群 1 和机群 3 内的机组相比,机群

2 和机群 4 内机组的有功功率明显较低。而从第 5 小时的数据分布情况可以明显地看出机群 1 与机群 3 的区别,该时段内机群 1 内的机组有功功率均位于 600 kW 以上区域,而机群 3 内的机组则位于低于 600 kW 以下区域。对比机群 2 和机群 4 的有功功率分布图可见,前 3 h 内机群 2 内机组的有功功率大致在 800~1 400 kW 之间,机群 4 内机组在 600~1 200 kW 之间。由此可见,采用本文提出的方法根据实测运行数据进行机群划分是能够根据数据的分布情况进行聚类的,可以反映各机组的实际运行情况。

根据相同的方法基于实测无功功率数据进行机群划分,结果如表 2 所示。

表 2 基于实测无功功率数据的机群划分结果
Table 2 Result of wind turbine grouping based on measured reactive power

机群号	机群内包含的风机编号
1	1、13、15、21、22、23、25、28、30、31、32、33
2	3、5、7、11
3	2、6、9、10、16、17、18、19、20、24、26、27
4	4、8、12、14、29

对比表 1 和表 2 可以看出,基于实测有功功率和无功功率数据划分的结果相近。对于划分有争议的机组(13、22、23、29、31),先分别计算其在 2 个划分机群的质心(同群内所有机组实测运行数据样本组的平均值),再根据该机组与质心的欧氏距离确定该机组是按有功功率还是无功功率划分。最后机群划分结果如表 3 所示。

表 3 综合有功、无功实测运行数据的机群划分结果
Table 3 Result of wind turbine grouping based on measured operating data integrating active and reactive power

机群号	机群内包含的风机编号
1	1、13、15、21、25、28、30、31、32、33
2	3、5、7、11、29
3	2、6、9、10、16、17、18、19、20、22、23、24、26、27
4	4、8、12、14

从划分结果可以看出,由于风电场内的风速分布、地形地貌、尾流效应等诸多因素的影响,地理位置相隔较近的风电机组运行状况可能差别很大,划分在不同的机群内;而地理位置相隔较远的机组也可能运行状况相近,划分在同一机群内。可见,对于布局不规则的大型风电场的机群划分,不宜采用简单按地理位置的划分方法。

根据机群划分结果,可将风电场等值为 4 台双馈风电机组,其动态模型采用文献[23]中的双馈风电机组模型,根据容量加权法^[24]计算等值风电机组的参数。

2.2 风速扰动下的仿真分析

以国际大电网会议推荐的 CIGRE B4-39 风电场

并网系统为例,在 PSS/E 平台上进行仿真分析,该仿真系统如图 6 所示, B_9 为系统平衡节点,区域 1 为系统较坚强区域,区域 2 为风电并网区域,区域 3 为系统重负荷区域,将上述等值的 4 台风电机组并联接在风电场母线节点上。

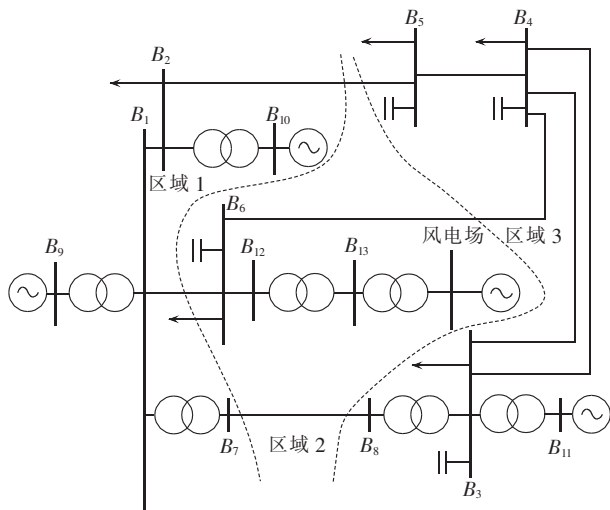


图 6 CIGRE B4-39 风电场并网仿真系统
Fig.6 Simulation system of CIGRE B4-39 grid-connected wind farm

以阵风扰动为例分析风速扰动下,风电场采用以下 4 种模型时的动态响应特性:传统的单机等值模型(模型 1);基于无监督谱聚类算法建立的多机表征模型(模型 2);基于半监督谱聚类算法建立的多机表征模型(模型 3);详细模型(模型 4)。假设迎向风电场的阵风 4 s 时启动,8 s 停止,阵风风速为 3 m/s,风电场有功功率动态响应曲线如图 7 所示。

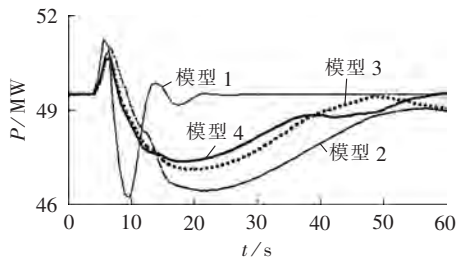


图 7 阵风扰动时风电场不同等值模型的有功出力
Fig.7 Active power output of different equivalent models for wind farm with gust disturbance

以详细模型结果为比较基准,表 4 给出仿真结果的平均绝对百分比误差(MAPE)和均方根误差(RMSE)。

表 4 阵风扰动下风电场采用不同等值模型时的误差比较

Table 4 Comparison of error among different equivalent models for wind farm with gust disturbance

等值模型	MAPE	RMSE
模型 1	0.034	1.862
模型 2	0.015	0.835
模型 3	0.006	0.369

可见,由于单机等值模型假设风电场内所有风电机组的运行状态完全一致,风电场采用单机等值模型时,风速扰动的影响被显著放大;采用多机表征模型减缓了整个风电场输出功率的波动;而半监督谱聚类算法引入了样本组的先验信息指导其后的聚类划分过程,明显提高了聚类效果,仿真结果比采用无监督谱聚类算法建立的多机表征模型能更准确地反映风电场的实际动态特性。

2.3 电网侧短路故障的仿真分析

设在 $t=2.0$ s 时,CIGRE B4-39 系统在 B_5 发生三相接地短路故障, $t=2.2$ s 时故障消除。风电场的有功出力、无功出力和机端电压标幺值的动态曲线分别如图 8—10 所示。

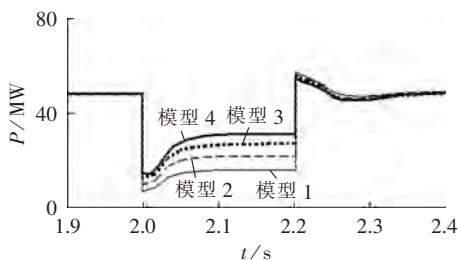


图 8 短路故障时风电场不同等值模型的有功出力
Fig.8 Active power output of different equivalent models for wind farm with short circuit fault

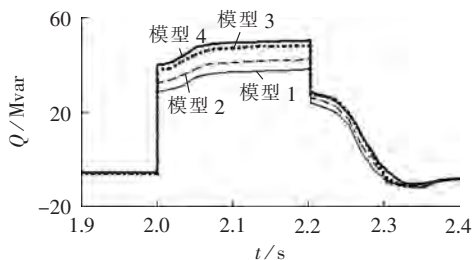


图 9 短路故障时风电场不同等值模型的无功出力
Fig.9 Reactive power output of different equivalent models for wind farm with short circuit fault

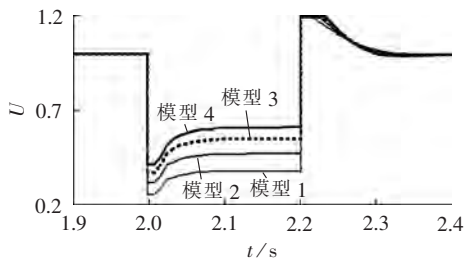


图 10 短路故障时风电场不同等值模型的机端电压
Fig.10 Terminal voltage of different equivalent models for wind farm with short circuit fault

由于短路点发生在电网侧,且 B_1 可以向并网点传输无功功率,风电场机端电压下降程度并不是很大,而由于故障发生在系统重负荷区域,受端电压跌落较大,需要提供大量的无功支持,因此故障期间风电场的无功出力大幅增加。

从上述仿真结果可以看出,与单机等值模型相比,多机表征模型考虑了机组风况差别、风电场内部各机组之间的相互作用等影响,在风速扰动和电网侧发生短路故障时,风电场的动态响应曲线与风电场详细模型更接近。而采用本文的半监督机群划分方法引入了样本组的先验信息指导其后的聚类划分过程,提高了机群聚类效果,相对于无监督谱聚类算法而言,划分结果更能反映风电机组的实际运行情况,建立的模型能够较准确地反映风电机组的轴系特性和转动惯量等,风电场的动态响应特性与详细模型更接近,模型的精确性有了一定的提高。

3 结论

本文面向风电场动态等值建模,提出一种基于分裂层次半监督谱聚类算法的风电场机群划分方法。采用该方法能够根据风电机组实测运行数据样本组的分布情况确定机群划分的个数,且在构建相似性矩阵时,采用自适应尺度参数代替固定的尺度参数,消除了尺度参数对聚类效果的影响。同时,在机群划分的过程中,利用获取的部分样本组的先验信息进行半监督性的学习,使同群内的机组相似性更高,提高了聚类效果。通过算例仿真可以看出,采用本文提出的方法根据实测运行数据进行机群划分是能够根据数据的分布情况进行聚类的,可以反映各机组的实际运行情况。本文建立的风电场动态模型在风速扰动和电网故障情况下均具有良好的适应性,动态响应特性与详细模型较接近,能够较准确地反映风电场的实际运行状况。

另外,从机群划分结果可以看出,对于机组布局不规则、地形复杂的风电场,当采用简单按地理位置的划分方法聚类时,会有较大的误差,如果模型精度要求高,则应按风电机组运行点相近的原则进行机群划分。

需要指出的是,机群划分的结果与实测运行数据样本选取的时间长度、时间段的代表性有关,经本文研究,选取典型、具有相近季节变化特征的实测运行数据进行建模,能有效提高模型的精确性。另外,如果实测运行数据样本空间中的样本组分布较分散,可能会使机群划分的个数较多,这种情况下如何考虑降低机群划分的个数,使建立的模型适用于实际的仿真分析也是一个值得深究的问题。对于以上问题,笔者将在后续工作中展开。

参考文献:

[1] 钱少锋,林俐,沈辉,等. 基于PSS/E Wind大型风电场并入输电网的动态特性研究[J]. 电力系统保护与控制,2009,37(6):11-16.
QIAN Shaofeng, LIN Li, SHEN Hui, et al. Dynamic characteristic analysis of transmission grid including large wind farm by PSS/E

wind package[J]. Power System Protection and Control,2009,37(6):11-16.
[2] 李洪美,万秋兰,向昌明. 考虑风速的风电场等值方法[J]. 电力自动化设备,2013,33(1):121-123,159.
LI Hongmei, WAN Qiulan, XIANG Changming. Wind farm equivalence method considering wind speed[J]. Electric Power Automation Equipment,2013,33(1):121-123,159.
[3] 温步瀛,江岳文,陈冲. 风电场并网运行的无功补偿优化问题[J]. 电力自动化设备,2008,28(5):42-46.
WEN Buying, JIANG Yuewen, CHEN Chong. Reactive power compensation optimization for grid-connected wind farm[J]. Electric Power Automation Equipment,2008,28(5):42-46.
[4] 曹娜,赵海翔,任普春,等. 风电场动态分析中风速模型的建立及应用[J]. 中国电机工程学报,2007,27(36):68-72.
CAO Na, ZHAO Haixiang, REN Puchun, et al. Establish and application of wind speed model in wind farm dynamic analysis[J]. Proceedings of the CSEE,2007,27(36):68-72.
[5] 陈树勇,戴慧珠,白晓民,等. 尾流效应对风电场输出功率的影响[J]. 中国电力,1998,31(11):28-31.
CHEN Shuyong, DAI Huizhu, BAI Xiaomin, et al. Impact of wind turbine wake on wind power output[J]. Electric Power,1998,31(11):28-31.
[6] 郑睿敏,李建华,李作红,等. 考虑尾流效应的风电场建模以及随机潮流计算[J]. 西安交通大学学报,2008,42(12):1515-1520.
ZHENG Ruimin, LI Jianhua, LI Zuohong, et al. Modeling of large-scale wind farms in the probabilistic power flow analysis considering wake effects[J]. Journal of Xi'an Jiaotong University,2008,42(12):1515-1520.
[7] 苏勋文,赵振兵,陈盈今,等. 尾流效应和时滞对风电场输出特性的影响[J]. 电测与仪表,2010,47(3):28-31.
SU Xunwen, ZHAO Zhenbing, CHEN Yingjin, et al. Impacts of wake effect and time delay on output characteristics of wind farm[J]. Electrical Measurement & Instrumentation,2010,47(3):28-31.
[8] 陈迎. 基于运行数据的风电场等效建模研究[D]. 北京:华北电力大学,2012.
CHEN Ying. Research on wind farm equivalent modeling based on operation data[D]. Beijing:North China Electric Power University,2012.
[9] 米增强,苏勋文,杨奇逊,等. 风电场动态等值模型的多机表征方法[J]. 电工技术学报,2010,25(5):162-169.
MI Zengqiang, SU Xunwen, YANG Qixun, et al. Multi-machine representation method for dynamic equivalent model of wind farms[J]. Transactions of China Electrotechnical Society,2010,25(5):162-169.
[10] 陈树勇,王聪,申洪,等. 基于聚类算法的风电场动态等值[J]. 中国电机工程学报,2012,32(4):11-19.
CHEN Shuyong, WANG Cong, SHEN Hong, et al. Dynamic equivalence for wind farms based on clustering algorithm[J]. Proceedings of the CSEE,2012,32(4):11-19.
[11] 齐雯. 大型风电场等值建模及其并网稳定性研究[D]. 北京:北京交通大学,2013.
QI Wen. Equivalent modeling and grid-connected stability of large wind farms[D]. Beijing:Beijing Jiaotong University,2013.
[12] 王聪. 基于聚类算法的风电场动态等值问题的研究[D]. 吉林:东北电力大学,2012.
WANG Cong. The dynamic equivalence for wind farm based on clustering algorithm[D]. Jilin:Northeast Dianli University,2012.

- [13] 高峰,赵东来,周孝信,等. 直驱式风电机组风电场动态等值[J]. 电网技术,2012,36(12):222-227.
GAO Feng,ZHAO Donglai,ZHOU Xiaoxin,et al. Dynamic equivalent algorithm for wind farm composed of direct-drive wind turbines[J]. Power System Technology,2012,36(12):222-227.
- [14] 蔡晓妍,戴冠中,杨黎斌. 谱聚类算法综述[J]. 计算机科学,2008,35(17):14-18.
CAI Xiaoyan,DAI Guanzhong,YANG Libin. Survey on spectral clustering algorithms[J]. Computer Science,2008,35(17):14-18.
- [15] 林俐,陈迎. 基于扩散映射理论的谱聚类算法的风电场机群划分[J]. 电力自动化设备,2013,33(6):113-118.
LIN Li,CHEN Ying. Wind turbine grouping with spectral clustering algorithm based on diffusion mapping theory [J]. Electric Power Automation Equipment,2013,33(6):113-118.
- [16] 沈亚田,沈夏炯,张磊. 基于图划分的谱聚类算法在文本挖掘中应用[J]. 计算机技术与发展,2009,19(5):96-98.
SHEN Yatian,SHEN Xiajiong,ZHANG Lei. Application of spectral cluster algorithm based on graph partition in text mining [J]. Computer Technology and Development,2009,19(5):96-98.
- [17] HERTZ T,SHENTAL N,BAR-HILLEL A,et al. Enhancing image and video retrieval;learning via equivalence constraint [C]// Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. Madison,USA:IEEE Computer Society,2003:668-674.
- [18] 赵凤焦,李成,刘汉强,等. 半监督谱聚类特征向量选择算法[J]. 模式识别与人工智能,2011,24(1):48-56.
ZHAO Fengjiao,LI Cheng,LIU Hanqiang,et al. Semi-supervised eigenvector selection for spectral clustering[J]. Pattern Recognition and Artificial Intelligence,2011,24(1):48-56.
- [19] 卜德云,张道强. 自适应谱聚类算法研究[J]. 山东大学学报:工学版,2009,39(5):22-26.
BU Deyun,ZHANG Daoqiang. Adaptive spectral clustering algorithm[J]. Journal of Shandong University:Engineering Science,2009,39(5):22-26.
- [20] 田铮,李小斌,句彦伟. 谱聚类的扰动分析[J]. 中国科学,2007,37(4):527-543.
TIAN Zheng,LI Xiaobin,JU Yanwei. Spectral clustering based on matrix perturbation theory[J]. Science in China,2007,37(4):527-543.
- [21] 孔万增,孙志海,杨灿,等. 基于本征间隙与正交特征向量的自动谱聚类[J]. 电子学报,2010,38(8):1880-1885.
KONG Wanzeng,SUN Zhihai,YANG Can,et al. Automatic spectral clustering based on eigengap and orthogonal eigenvector [J]. Acta Electronica Sinica,2010,38(8):1880-1885.
- [22] HARB A M,HARB B A. Controlling chaos in Josephson junction using nonlinear backstepping controller[J]. IEEE Transactions on Applied Superconductivity,2006,16(4):1988-1998.
- [23] 徐琼璟,徐政. PSS/E 中的风电机组通用模型概述[J]. 电网技术,2010,34(8):176-182.
XU Qiongjing,XU Zheng. A survey on general models of wind turbine generators in PSS/E[J]. Power System Technology,2010,34(8):176-182.
- [24] 胡杰,余贻鑫. 电力系统动态等值参数聚合的实用方法[J]. 电网技术,2006,30(24):26-30.
HU Jie,YU Yixin. A practical method of parameter aggregation for power system dynamic equivalence[J]. Power System Technology,2006,30(24):26-30.

作者简介:



林俐(1968—),女,云南会泽人,副教授,博士,主要从事电力系统分析与控制、风电并网及其系统分析方面的研究工作(E-mail:linli@ncepu.edu.cn)。

Wind turbine grouping based on semi-supervised split-hierarchical spectral clustering algorithm for wind farm

LIN Li,PAN Xianxian

(State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources,
North China Electric Power University,Beijing 102206,China)

Abstract: A method of wind turbine grouping based on semi-supervised split-hierarchical algorithm is proposed,which builds the dynamically equivalent model of wind farm based on the spectral clustering technology. An eigenvector matrix Y is formed based on the measured operating data of wind turbines, which reflects the original data structure and provides more effective information for clustering. With the available prior information of some samples,the semi-supervised split-hierarchical spectral clustering strategy is adopted to divide the sample groups of Y into several clusters as the wind turbine groups. The capacity weighting method is applied to calculate the parameters of each equivalent wind turbine group and build the dynamically equivalent model of wind farm. As an example,results of the simulation for an actual wind farm show that,the dynamically equivalent model built approximates to the detailed model,accurately reflecting the dynamic response characteristics of wind farm.

Key words: split hierarchy; semi-supervision; spectral clustering; wind farms; wind turbine grouping; clustering algorithms; dynamic equivalence