

基于扩散映射理论的谱聚类算法的风电场机群划分

林 俐¹, 陈 迎^{1,2}

(1. 华北电力大学 新能源电力系统国家重点实验室, 北京 102206;

2. 安徽省电力公司检修公司, 安徽 合肥 230000)

摘要: 针对地形复杂或布局不规则的风电场, 将谱聚类方法应用于风电场机群划分, 提出了一种风电场的机群分类方法。该方法以风电机组具有相同或相近运行点为机组分群原则, 应用基于扩散映射理论的谱聚类算法对风电场各机组的实测运行数据进行聚类分析, 找到风电机组之间动态运行过程的相似性, 从而实现对风电场内所有风电机组的聚类划分。通过算例仿真验证了所提出的机群划分方法的有效性。

关键词: 风电场; 动态等值; 机群划分; 谱聚类; 聚类算法

中图分类号: TM 862; TM 614

文献标识码: A

DOI: 10.3969/j.issn.1006-6047.2013.06.021

0 引言

风力发电在我国迅速发展。截至 2010 年底, 中国全年风力发电新增装机容量达 18 928 MW, 累计装机容量达到 44 733 MW, 占全球风电总装机容量的 22.7%^[1], 居世界第一。

大型风电场通常由数十台甚至上百台风力发电机组组成, 每台风电机组都是一个非线性的动态系统。在风电场并网的电力系统仿真分析中, 若对每台风电机组进行详细建模将大幅增加风电场模型的复杂度, 并且计算时间长、占用内存大。因此, 有必要对整个风电场进行等值, 利用等值模型进行风电场并网的分析和计算。

目前通常是采用单机等值一个风电场^[2-7]。然而, 对于大型的风电场, 由于地形地貌以及尾流效应和时滞的影响, 风电场内风速分布不均匀, 风电机组的风速差异较大, 使用单机等值方法通常会存在较大误差^[8]。

将风电场内所有风电机组进行机群划分, 再对同群的机组进行合并等值处理, 是简化风电场模型的有效方法。文献[9-10]指出风力发电机组的运行情况主要由风速决定, 而尾流效应是造成风电场不同区域的风速差异, 使得风力发电机组的运行点不同的主要原因, 进而提出依据尾流效应对风电场机组进行划分, 认为在相同风速扰动下的风电机组群具有相同或接近的动态过程。文献[11]指出对定速风电机组而言, 转速比风速更能准确地反映风电机组的运行点, 并针对风速差异较大的定速机组风电场, 定义了一种以故障切除时风电机组的转速为机群分

类指标的新的机群划分方法。

上述 2 种机群分类方法都是以风电机组具有相同或相近运行点作为机组分群原则, 分别以风速和转速反映风电机组的运行点。本文考虑到各风电机组的有功、无功出力和电压这些重要的输出特性, 综合了风电机组的运行环境、所受风速的波动、转速和风电机组参数等运行信息(包含风电机组在风速波动时的动态信息), 可以直接、准确地反映风电机组的运行点。因此, 可以应用工程上实用的统计综合法, 由风电机组运行情况的信息对特性相似的风电机组进行聚类划分。

对于相同类型风电机组组成的风电场, 机群划分的方法与风电场内风机的布局以及风电场所处的地形有着密切的关系。对于选址在平地或海上的风电场, 风机通常是在主风方向上按照行列规则分布的, 这种情况下按照机组的安装位置, 将分布在主风方向上的同排风机划分到同一机群的机群划分方法是合理的。但是, 对于地形复杂或布局不规则的风电场, 由于每台风机所处的地形、海拔高度、布局等因素存在较大差异, 每台风机的来风将有较大不同, 即使是同排风机运行的工作点也可能有较大差异。

本文从电力系统仿真的角度, 针对地形复杂或布局不规则的风电场, 将谱聚类方法应用于机群划分, 提出了一种风电场的机群分类方法。该方法以风电机组具有相同或相近运行点为机组分群原则, 应用基于扩散映射理论的谱聚类算法对风电场各机组的实测运行数据进行聚类分析, 找到风电机组之间动态运行过程的相似性, 从而实现对风电场内所有风电机组的聚类划分。以国内某双馈风力发电机组组成的风电场为例, 基于风电场实测的运行数据对风电场装设的 23 台风机进行了聚类划分, 建立了风电场等值模型。最后, 在 PSS/E 仿真平台上进行算例仿真, 并与风电场详细模型进行比较分析, 验证了该方法的有效性。

收稿日期: 2012-06-06; 修回日期: 2013-04-24

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

Project supported by the National Sci-Tech Support Plan (2013BAA02B01) and the Key Program of National Natural Science Foundation of China(51190103)

1 基于扩散映射理论的谱聚类分析方法

所谓聚类,就是将数据对象划分成几个不同的类别,使在同一类中的数据对象尽可能相似,在不同类中的数据对象差别尽可能大,聚类分析是一种广泛应用于数据挖掘和数据分析的有效方法^[12]。近年来所提出的谱聚类是一种较为实用的聚类方法^[13-14],目前已出现多种谱聚类算法,如 Shi 和 Malik 提出的 Ncut 算法^[15]和 Ng 等提出的 NJW 算法^[16],首先利用数据样本构造出一个相似性矩阵,进而计算得出拉普拉斯矩阵,然后利用拉普拉斯矩阵的特征向量来找出数据样本的内在联系。2001 年,Meila 和 Shi 将两点间的相似度解释为马尔可夫链中随机游走的概率,并使用马尔可夫概率转移矩阵的特征向量进行聚类,从另一个角度解释了谱聚类^[17]。

本文所采用的谱聚类方法包括以下几个步骤:

a. 基于风电场各机组的实测数据样本,构建一个带高斯权重值的图来表示各机组之间数据的相似性,在该数据图上定义一个马尔可夫随机游走^[18-19]并构建马尔可夫概率转移矩阵,以概率的形式来表示数据点之间的相似度;

b. 对马尔可夫转移矩阵进行谱分解和特征降维来识别聚类类数;

c. 在数据点之间随机游走的关系上,定义扩散距离^[20-22]作为数据点之间接近度的度量,通过比较数据点两两之间的扩散距离和给定的阈值,将每个数据都聚类到相应的群,从而确定各个群具体包含哪些数据点,这样就把每台风电机组划分到相应的群。

1.1 马尔可夫转移矩阵的构建

如果风电场内有 n 台风机,设待聚类划分的样本集 $X = \{X_1, X_2, \dots, X_i, \dots, X_n\} \in \mathbb{R}^{m \times n}$, 其中 X_i 为待聚类的对象,是由第 i 台风机的采样数据 $x_i(t)$ 集合构成的一个高维数据点, $X_i = \{x_i(1), x_i(2), \dots, x_i(t), \dots, x_i(m)\}^T, t \in T$ 。由数据样本集 X 构造的带高斯权重的图 G ,如图 1 所示。 G 是个连通图,图中顶点表示数据点,顶点之间连线上的权重值表示数据点之间的相似性。相应地,可以得到图的权矩阵即相似性矩

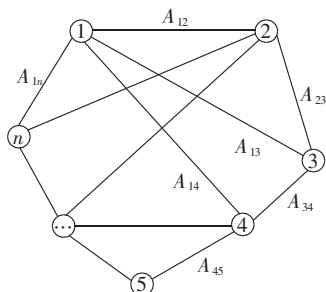


图 1 由数据样本集 X 构造的带高斯权重的图 G

Fig.1 Graph G with Gaussian weights constructed from data sample set X

阵 $A \in \mathbb{R}^{n \times n}$ 如下:

$$A_{ij} = e^{-\|x_i - x_j\|^2 / (2\sigma^2)} \quad (1)$$

其中, $\|x_i - x_j\|$ 表示两点之间的欧氏距离; σ 为尺度参数^[16],用来控制数据点 x_i 和 x_j 之间的距离对权值 A_{ij} 的影响,使相似矩阵的值都控制在较为相似的范围内。

设数据点 X_i 为一个具有马尔可夫链性质的随机过程向量,根据相似矩阵 A 定义随机过程样本集 X 的马尔可夫转移矩阵 $P \in \mathbb{R}^{n \times n}$ 如下:

$$P_{ij} = A_{ij} / \sum_{j=1}^n A_{ij} \quad (2)$$

由式(1)和(2)可知,矩阵 A 是对称的,矩阵 P 是可逆的。马尔可夫转移矩阵体现任意 2 台风机划入同一机群的概率。

1.2 马尔可夫概率转移矩阵的谱分析

设 λ_i, φ_i 和 ψ_i 分别表示矩阵 P 的第 i 个特征值及其对应的左特征向量和右特征向量。将矩阵 P 的特征值从大到小顺序排列,且 $1 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_i \geq \dots \geq \lambda_n \geq 0$, 相应的左特征向量和右特征向量分别为 $\varphi_1, \varphi_2, \varphi_3, \dots, \varphi_i, \dots, \varphi_n$ 和 $\psi_1, \psi_2, \psi_3, \dots, \psi_i, \dots, \psi_n$, 则马尔可夫概率转移矩阵 P 的谱分解形式如下:

$$P = \sum_{i=1}^n \lambda_i \psi_i \varphi_i^T \quad (3)$$

对矩阵 P 作特征降维^[23]处理:

a. 若矩阵 P 拥有 $q < n$ 个主导特征值,即 $\lambda_1, \lambda_2, \dots, \lambda_q$ 这 q 个特征值大小相差不大且接近于 1, 且 $\lambda_{q+1} \ll \lambda_q$, 则可以用前 q 个特征值和特征向量形成矩阵 P 的秩为 q 的近似矩阵 P_q ;

b. 若矩阵 P 的特征值不满足条件 a, 观察矩阵 P 相邻特征值的差值,若存在某个突然变小的特征值 λ_{q+1} , 即 $\lambda_q - \lambda_{q+1} \gg \lambda_k - \lambda_{k+1}$, 其中 $q+1 \leq k \leq n$, 且 λ_{q+1} 的值很小, 则同样可取前 q 个特征值和特征向量形成矩阵 P 的秩为 q 的近似矩阵 P_q 。

则矩阵 P 的秩为 q 的近似矩阵 P_q 的谱分解形式如下:

$$P_q = \sum_{i=1}^q \lambda_i \psi_i \varphi_i^T \quad (4)$$

矩阵 P 的特征向量空间的维数约减为 q , 则待聚类的对象集的聚类类数为 q , 即风电场的风机分群数为 q 。

1.3 扩散距离的计算

定义映射 $\psi: x \rightarrow \begin{bmatrix} \lambda_1 x \psi_1 \\ \vdots \\ \lambda_n x \psi_n \end{bmatrix}$, ψ 仍然是在 $\mathbb{R}^n \rightarrow \mathbb{R}^n$ 。

设任意 2 台风机 x_i 和 x_j 的扩散距离:

$$D^2(x_i, x_j) = D^2(e_i^T, e_j^T) = \|\psi(e_i^T) - \psi(e_j^T)\|^2 = \left\| \begin{bmatrix} \lambda_1 [\psi_1(i) - \psi_1(j)] \\ \vdots \\ \lambda_n [\psi_n(i) - \psi_n(j)] \end{bmatrix} \right\|^2 \quad (5)$$

其中, $\mathbf{e}_1, \dots, \mathbf{e}_n$ 为对应于 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 的单位向量。
取前 q 个主导特征值和相应的特征向量, 可以得到 $\mathbf{x}_i$ 和 $\mathbf{x}_j$ 的扩散距离的简化表达式:

$$D^2(\mathbf{x}_i, \mathbf{x}_j) \approx \left\| \begin{matrix} \lambda_1 [\boldsymbol{\psi}_1(i) - \boldsymbol{\psi}_1(j)] \\ \vdots \\ \lambda_q [\boldsymbol{\psi}_q(i) - \boldsymbol{\psi}_q(j)] \end{matrix} \right\|^2 \quad (6)$$

根据风电场风机的分群数以及风机之间扩散距离的分布情况, 选取阈值 $\eta > 0$ 。若 $D^2(\mathbf{x}_i, \mathbf{x}_j) \leq \eta$, 则表明第 i 台和第 j 台风机应该划入同一个机群; 若 $D^2(\mathbf{x}_i, \mathbf{x}_j) > \eta$, 则表明第 i 台和第 j 台风机不属于同一机群。

根据实际运行经验, 在上述机群划分结果的基础上, 还应作以下综合修正: 类型、运行方式不相同的风机不应划入同一机群; 风况非常接近的风机应划入同一机群。

2 实例分析

本文所采用的算例数据来自于我国某实际风电场。该风电场装设有 23 台单机容量为 1.5 MW 的双馈风力发电机组, 总容量为 34.5 MW。本文对该风电场 2010 年 8 月 1 日至 2010 年 8 月 31 日的实测数据进行分析。

2.1 实测数据的处理

考虑相邻时刻数据的差值能消除测量设备自身带来的误差, 比某一时刻的绝对输出值更能反映风机运行的变化特性。差值定义如下:

$$\Delta y(t) = y(t) - y(t-1) \quad (7)$$

以风电场中编号为 WT3 的风力发电机为例, 图 2 和图 3 分别为 WT3 的有功功率 P_{WT3} 和相邻时刻有功功率的差值 ΔP_{WT3} 。

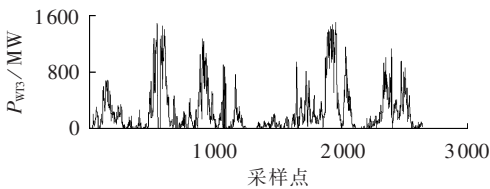


图 2 WT3 各采样点的有功功率
Fig.2 Active power of WT3 sample points

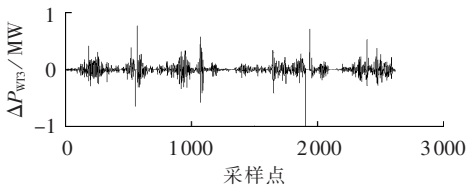


图 3 WT3 相邻采样点的有功功率差值
Fig.3 Active power difference between two adjacent sample points of WT3

2.2 建立马尔可夫概率转移矩阵

由式(1)和(2)建立风电场 23 台风机有功功率和无功功率的马尔可夫概率转移矩阵 $\mathbf{P}_P, \mathbf{P}_Q \in \mathbf{R}^{23 \times 23}$ 。

$$\mathbf{P}_P = \begin{bmatrix} 0.2244 & 0.0731 & \cdots & 0.0296 & 0.0176 \\ 0.0563 & 0.2917 & \cdots & 0.0253 & 0.0137 \\ \vdots & \vdots & & \vdots & \vdots \\ 0.0200 & 0.0222 & \cdots & 0.3315 & 0.0265 \\ 0.0116 & 0.0117 & \cdots & 0.0258 & 0.3400 \end{bmatrix}$$
$$\mathbf{P}_Q = \begin{bmatrix} 0.2948 & 0.0534 & \cdots & 0.0149 & 0.0076 \\ 0.0717 & 0.3954 & \cdots & 0.0165 & 0.0075 \\ \vdots & \vdots & & \vdots & \vdots \\ 0.0234 & 0.0192 & \cdots & 0.4610 & 0.0197 \\ 0.0117 & 0.0086 & \cdots & 0.0194 & 0.4545 \end{bmatrix}$$

2.3 马尔可夫概率转移矩阵的谱分析

求解马尔可夫概率转移矩阵的特征值和特征向量。 $\mathbf{P}_P$ 和 $\mathbf{P}_Q$ 的特征值计算结果如下:

$$\boldsymbol{\lambda}_P = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \dots, \lambda_{23}] = [1.0, 0.9055, 0.8081, 0.4187, 0.3729, 0.3482, 0.3359, \dots, 0.1139]$$
$$\boldsymbol{\lambda}_Q = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \dots, \lambda_{23}] = [1.0, 0.8125, 0.7370, 0.5548, 0.5105, 0.4974, 0.4859, \dots, 0.1686]$$

通过观察, $\mathbf{P}_P$ 的前 3 个特征值较大, 第 4 个特征值突然减小, 并且第 4 个之后的特征值之间的差值较小。根据 1.2 节所述的特征降维理论, 此风电场风电机组可以划分为 3 个机群。 $\mathbf{P}_Q$ 的特征值与 $\mathbf{P}_P$ 的特征值具有类似特点, 由此再次确定机群分类数为 3。

2.4 机群划分

本文取阈值 $\eta = 0.4$ 。根据有功功率和无功功率得到机群划分结果分别如表 1 和表 2 所示, 2 个机群划分结果基本相同。有划分争议的风电机组是 WT13、WT18。由有功功率概率转移矩阵计算出的风机间扩散距离显示, WT13、WT18 与 3 个机群的最近扩散距离分别为: 0.412 8, 0.367 5, 0.389 7; 0.434 4, 0.394 1, 0.371 3。而通过无功功率概率转移矩阵算出的 WT13、WT18 与 3 个机群的最近扩散距离分别为: 0.358 3, 0.417 4, 0.441 5; 0.360 9, 0.406 4, 0.413 2。

表 1 机群划分结果(根据有功功率划分)
Tab.1 Grouping results(according to active power)

对象		参数		
风机编号		1~23		
主导特征值个数		3		
机群编号	1 号机群	2 号机群	3 号机群	
同调群	1, 3, 7, 9,	2, 6, 11, 13,	4, 5, 8, 10, 12,	
风机编号	14, 17, 19, 21	16, 23	15, 18, 20, 22	

表 2 机群划分结果(根据无功功率划分)
Tab.2 Grouping results(according to reactive power)

对象		参数		
风机编号		1~23		
主导特征值个数		3		
机群编号	1 号机群	2 号机群	3 号机群	
同调群	1, 3, 7, 9, 13, 18,	2, 6, 11,	4, 5, 8, 10, 12,	
风机编号	14, 17, 19, 21	16, 23	15, 20, 22	

取这 2 种扩散距离的均值,则风机 WT13、WT18 与 3 个机群的扩散距离分别为:0.385 5,0.392 4,0.415 6;0.397 7,0.400 2,0.392 2。这样 WT13 应划入 1 号机群,WT18 应划分到 3 号机群。从而得到如表 3 所示的机群最终划分结果。

表 3 机群最终划分结果

Tab.3 Final results of wind turbine grouping

对象	参数		
风机编号	1~23		
主导特征值个数	3		
机群编号	1 号机群	2 号机群	3 号机群
同调群	1,3,7,9,13,	2,6,11,	4,5,8,10,12,
风机编号	14,17,19,21	16,23	15,18,20,22

将表 3 所示的机群划分结果与风电场的实际情况对比分析发现:

a. 本风电场所处地形比较复杂,机群的划分结果与按地理位置的划分结果基本吻合,但又不完全相同,如风机 WT14 与 WT2、WT6、WT11、WT16 按照一条直线分布在主风向上,但由于 WT14 的海拔高度和其他风机差距较大,最终没有和同排的其他风机划分在同一个机群内;

b. 受到地形、尾流效应等因素的影响,有些风机之间地理距离较近,但是却不在一个机群内,如风机 WT23 与 WT10、WT12 地理上的距离较近,海拔高度接近,但受到主风向上游的 WT9 风机的尾流效应影响,WT23 没有与 WT10、WT12 划入同一机群。

对比分析说明,对于地形复杂、布局不规则的风电场,采用按地理位置简单划分和按风电机组运行点相近的原则划分的结果往往不同。

3 仿真验证

3.1 风电场仿真系统及电气参数

选取国际大电网会议推荐的 CIGRE B4-39 风电场并网系统,如图 4 所示,基于 PSS/E 仿真软件进行仿真分析。

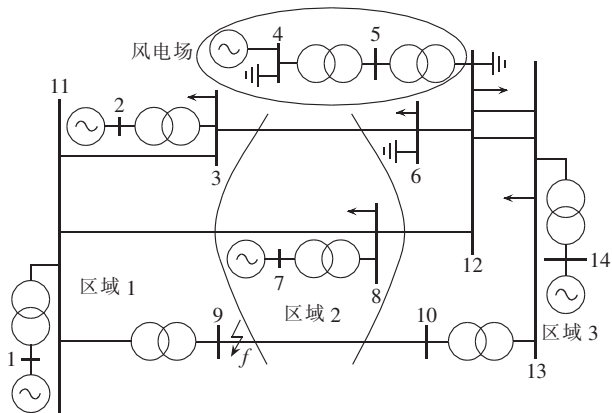


图 4 CIGRE B4-39 风电场并网系统

Fig.4 CIGRE B4-39 grid-connected wind farm

在母线 4 上并入 23 台单机容量为 1.5 MW 的双馈风电机组,该风电场通过 2 级变压器升压至 230 kV。根据 2.4 节的机群划分结果,将整个风电场等值为 3 台风电机组。由于该风电场的风电机组都为同型号的双馈风电机组,采用文献[24]的容量加权法计算等值风电机组的参数。

3.2 风电场详细模型与等值模型的仿真比较

分别针对电网侧短路故障和风速波动 2 种情况进行仿真验证。仿真曲线图中 A 表示风电场等值模型,B 表示风电场详细模型;电压为标幺值。

3.2.1 电网侧短路故障仿真

设定在 $t = 1.0\text{ s}$ 时,CIGRE B4-39 系统母线 9 发生三相接地短路故障, $t = 1.12\text{ s}$ 时故障消除。图 5 给出了仿真结果。

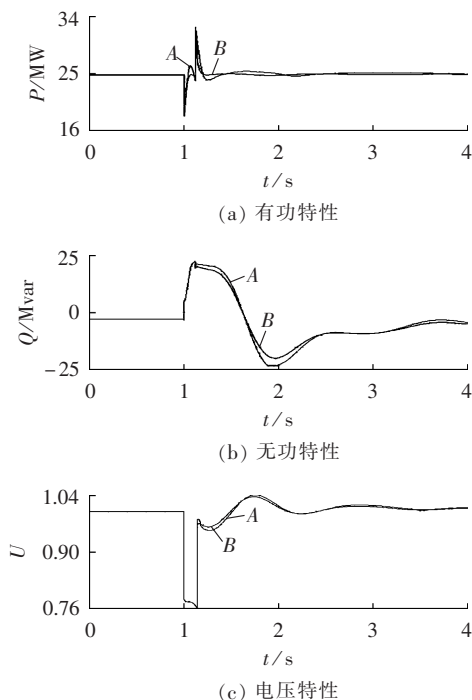


图 5 短路故障时风电场的动态响应

Fig.5 Dynamic response of wind farm to short circuit fault

3.2.2 风速波动仿真

由于随机风变化较快,变化幅度不大,加之电力系统的惯性较大,随机风对电力系统的动态特性产生的影响较小,因此在仿真时可以忽略随机风的影响。图 6(a) — (c) 给出了 0~50 s 内风电场并网点动态仿真结果,仿真所采用的风速 v 波动曲线如图 6 (d) 所示。

3.2.3 仿真分析

以动态响应的最大偏差和恢复稳定的时间作为 2 个指标,分别定义如下:

$$\delta_{\max} = \max_{k=0}^K \frac{|R(k) - R_f(k)|}{R_f(k)} \quad (8)$$

$$\tau = |t - t_f| / t_f \quad (9)$$

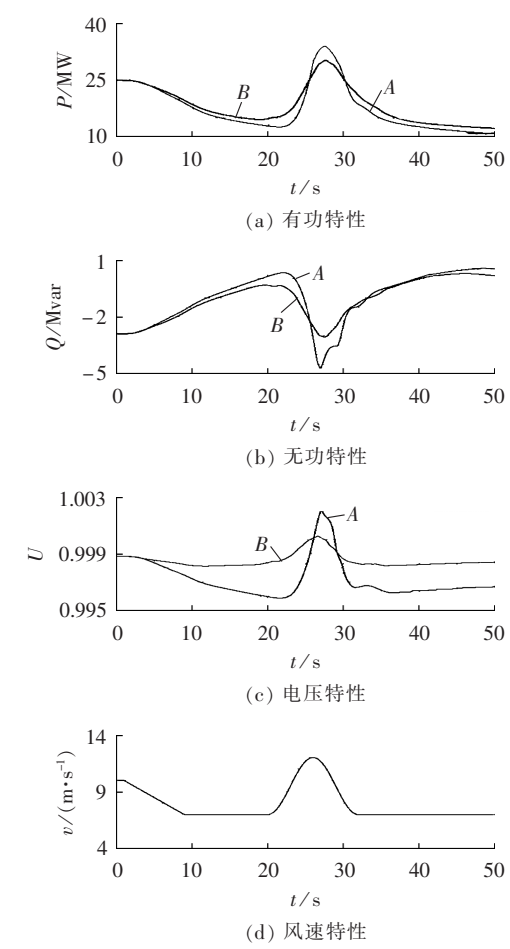


图 6 风速波动下风电场的动态响应
Fig.6 Dynamic response of wind farm to wind speed fluctuation

其中, R 和 R_f 分别为风电场等值模型和风电场详细模型的动态响应; K 为动态响应按时间采样的点数; t 和 t_f 分别为采用风电场等值模型和风电场详细模型时扰动后恢复稳定的时间。

电网侧短路故障和风速波动 2 种动态过程的风电场等值模型和风电场详细模型动态响应偏差如表 4 和表 5 所示。

表 4 电网侧短路故障时风电场等值模型与详细模型动态响应偏差

Tab.4 Difference in dynamic response to short circuit fault between equivalent model and detailed model

对象	$\delta_{\max}$	τ
有功功率	2.1%	1.3%
无功功率	4.2%	0.9%
电压	0.023%	0.17%

从动态响应的最大偏差和恢复稳定的时间 2 个指标来看,风电场等值模型和详细模型动态响应差距不大,对于电网侧短路故障和风速波动 2 种动态过程都有较好的精度。这证明了本文的机群划分方法对于电力系统仿真是有效的。

表 5 风速波动时风电场等值模型与详细模型动态响应偏差

Tab.5 Difference in dynamic response to wind speed fluctuation between equivalent model and detailed model

对象	$\delta_{\max}$	τ
有功功率	9.5%	2.2%
无功功率	7.6%	1.4%
电压	0.2%	1.8%

4 结论

本文从电力系统仿真的角度,将谱聚类方法应用到机群划分中,提出了一种适用于地形复杂或布局不规则的风电场的机群分类方法。该方法以风电机组具有相同或相近运行点为机组分群原则,根据风电场各机组的实测运行数据,应用基于扩散映射理论的谱聚类算法对风电机组进行分群。基于本文所提出的机群分类方法建立的风电场等值模型能够较准确地反映风电场并网点的动态特性,适用于分析大容量风电场对电力系统的影响,具有重要的工程应用价值。另外,该方法对于恒速风力发电机风电场和变速风力发电机风电场的动态等值都适用。

参考文献:

[1] Global Wind Energy Council. GWEC global wind report 2010[R]. Belgium, Brussels; [s.n.], 2011.

[2] 乔嘉庚,鲁宗相,闵勇,等. 风电场并网的新型使用等效方法[J]. 电工技术学报,2009,24(4):209-213.

QIAO Jiageng, LU Zongxiang, MIN Yong, et al. New dynamic equivalence method for grid-connected wind farm[J]. Transactions of China Electrotechnical Society, 2009, 24(4): 209-213.

[3] PERDANA A, USKI-JOUTSENVUO S, CARLSON O, et al. Comparison of an aggregated model of a wind farm consisting of fixed-speed wind turbines with field measurement [J]. Wind Energy, 2008, 11(1): 13-27.

[4] 王磊,贾宏杰. 无功协调控制策略对风电场并网系统小扰动稳定性的影响[J]. 电力自动化设备, 2012, 32(11): 28-33.

WANG Lei, JIA Hongjie. Impact of coordinated var control on small signal stability of power system with wind farm[J]. Electric Power Automation Equipment, 2012, 32(11): 28-33.

[5] SLOOTWEG J G, KLING W L. Aggregated modelling of wind parks in power system dynamics simulations[C]// 2003 IEEE Power Technology Conference. Bologna, Italy: IEEE, 2003: 626-631.

[6] SHAFIU A, ANAYA-LARA O, BATHURST G. Aggregated wind turbine models for power system dynamic studies[J]. Wind Engineering, 2006, 30(3): 171-186.

[7] 金海峰,吴涛. 风电接入系统后的电压稳定问题[J]. 电力自动化设备, 2010, 30(9): 82-84.

JIN Haifeng, WU Tao. Voltage stability after grid-connection of wind farm[J]. Electric Power Automation Equipment, 2010, 30(9): 82-84.

[8] PERDANA A. Dynamic models of wind turbines: a contribution towards the establishment of standardized models of wind turbines for power system stability studies [D]. Göteborg, Sweden: Chalmers University of Technology, 2008.

- [9] AKHMATOV V, KNUDSEN H. An aggregate model of a grid-connected, large-scale, offshore wind farm for power stability investigations: importance of windmill mechanical system[J]. *Electrical Power Energy Systems*, 2002(24):709-717.
- [10] FERNÁNDEZ L M, SAENZ J R, JURADO F. Dynamic models of wind farms with fixed speed wind turbines[J]. *Renewable Energy*, 2006, 31(8):1203-1230.
- [11] 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):162-167.
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):162-167.
- [12] JAIN A, MURTY M, FLYNN P. Data clustering: a review [J]. *ACM Computing Survey*, 1999, 31(3):264-323.
- [13] KANNAN R, VEMPALA S, VETTA A. On clusterings: good, bad, and spectral[J]. *Journal of the ACM*, 2004, 51(3):497-515.
- [14] LUXBURG U. A tutorial on spectral clustering[J]. *Statistics and Computing*, 2007, 17(4):395-416.
- [15] SHI J, MALIK J. Normalized cuts and Image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8):888-905.
- [16] NG A Y, JORDAN M I, WEISS Y. On spectral clustering: analysis and an algorithm[C]//Proc of NIPS 2001. Vancouver, Canada: [s.n.], 2001:849-856.
- [17] MEILA M, SHI J. A random walks view of spectral segmentation[C]//Proc of Eighth International Conference on AI and Statistics. Key West, Florida, USA: [s.n.], 2001:873-879.
- [18] RUNOLFSSON T, MA Y. Model reduction of nonreversible Markov chains[C]//Proceedings of the IEEE CDC. New Orleans, USA: IEEE, 2007:3739-3744.
- [19] SZUMMER M, JAAKKOLA T. Partially labeled classification with Markov random walks[C]//Neural Information Processing Systems. Vancouver, Canada: [s.n.], 2001:945-952.
- [20] NADLER B, LAFON S, COIFMAN R R, et al. Diffusion maps, spectral clustering and the reaction coordinate of dynamical systems[J]. *Applied and Computation Harmonic Analysis: Special Issue on Diffusion Maps and Wavelets*, 2006, 21(1):113-127.
- [21] COIFMAN R R, KEVREKIDIS I G, LAFON S, et al. Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems[J]. *Multiscale Model and Simulation*, 2008, 7(2):842-864.
- [22] MAGGIONI M, MHASKAR H. Diffusion polynomial frames on metric measure spaces[J]. *Applied and Computation Harmonic Analysis*, 2008, 24(3):329-353.
- [23] DEUHARD A F P, HUISINGA W, SCHÜTTE C. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chain[J]. *Linear Algebra and its Applications*, 2000, 315(1):39-59.
- [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(4):26-30.
HU Jie, YU Yixin. A practical method of parameter aggregation for power system dynamic equivalence[J]. *Power System Technology*, 2006, 30(4):26-30.

作者简介:

林 俐(1968-), 女, 云南会泽人, 副教授, 博士, 主要研究方向为电力系统分析、运行与控制及风电并网分析(E-mail: linli@ncepu.edu.cn);

陈 迎(1987-), 女, 湖北仙桃人, 助理工程师, 硕士, 主要从事电力系统分析、运行与控制方面的工作。

Wind turbine grouping with spectral clustering algorithm based on diffusion mapping theory

LIN Li¹, CHEN Ying^{1,2}

(1. State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources,

North China Electric Power University, Beijing 102206, China;

2. Anhui Electric Power Corporation, Hefei 230000, China)

Abstract: For the wind farms with complex terrain or irregular layout, a wind turbine aggregation method based on the spectral clustering technique is proposed, which groups the wind turbines with same or similar operating point together. The spectral clustering algorithm based on the diffusion mapping theory is employed, which clusters the measured operating data of all wind turbines in a wind farm, captures the similarity among them in the dynamic operating process, and divides them into different groups. Simulative verification demonstrates the effectiveness of the proposed grouping method.

Key words: wind farms; dynamic equivalence; wind turbine grouping; spectral clustering; clustering algorithms