

## 基于膨胀腐蚀聚类方法的风电功率预测

周 晓<sup>1</sup>, 冬 雷<sup>1</sup>, 郝 颖<sup>1</sup>, 廖晓钟<sup>1</sup>, 高 阳<sup>2</sup>

(1. 北京理工大学自动化学院, 北京 100081; 2. 沈阳工程学院电力学院, 沈阳 110136)

**摘 要:** 提出一种基于膨胀腐蚀的聚类方法, 并利用 UCI(university of california irvine)数据集进行实验仿真证明此方法的可行性。将此聚类方法应用于风电功率预测中的 NWP(numerical weather prediction)信息分类, 选择与预测日同一类的历史日数据作为训练样本, 利用广义回归神经网络预测功率, 并与直接预测的方法相比较, 仿真结果表示基于膨胀腐蚀对历史日数据分类后再预测的精度较高。

**关键词:** 风电功率预测; 聚类; 膨胀腐蚀; 广义回归神经网络

**中图分类号:** TK614

**文献标识码:** A

### 0 引 言

风力发电机组实际运行中产生功率的大小主要取决于当地气象资源的情况, 尤其是风速的变化特性。风速主要受气象因素及地形、地表障碍物等地形因素的影响。地球自转产生的昼夜交替使得某些天的天气状况呈现某种程度的相似性, 所以不同日的风功率变化趋势十分近似。风电功率变化趋势近似的日子存在某些共有的特征, 体现在数值天气预报(numerical weather prediction NWP)信息变化趋势基本一致。因此对历史 NWP 气象样本数据进行聚类后, 提取与预测日 NWP 信息相似性高的历史样本作为预测模型的训练输入, 可在提高风电功率预测精度的同时大大减小建模仿真的计算量<sup>[1]</sup>。

聚类是指在不知道任何先验知识的条件下, 根据数据之间的相似性和距离将数据集分为几类的过程。因为聚类的数量和结构均未事先假定, 所以聚类分析是一种无监督的学习方法<sup>[2]</sup>。聚类作为一种重要的数据分析和挖掘手段, 已广泛应用于许多领域<sup>[3,4]</sup>。迄今为止, 现有的聚类算法主要分为基于层次的聚类、基于划分的聚类、基于密度的聚类、基于网格的聚类、基于模型的聚类等<sup>[3]</sup>。然而这些算法均需要事先指定聚类数量  $k$  的值, 因此确定聚类数量成为需要解决的难题。目前确定  $k$  值的方法大

致可以分为 3 类: 1) 设置一个聚类有效性指标, 利用合并或者分裂规则增加或者减少  $k$  直到达到聚类最优<sup>[5-8]</sup>。2) 直接求  $k$  值。孙昌思核等<sup>[9]</sup>通过数据亲和矩阵的谱分解得到一系列特征值, 再利用特征值差值分析来确定类数量。文献<sup>[10]</sup>通过 NNS 法(neighbor sharing selection)识别高密度区来确定聚类数量和初始聚类中心。谢娟英等<sup>[11]</sup>通过构造样本距离相对于样本密度的决策图来确定聚类数量。3) 在聚类的时候自动确定聚类数量<sup>[12-14]</sup>。

本文提出一种基于膨胀腐蚀的聚类方法, 首先数据经过处理后对应一个二值矩阵, 再对该二值矩阵进行膨胀腐蚀得到新的二值矩阵, 提取新的二值矩阵中相连的值为 1 的数据点为一类, 包含数据量相对较多的类的数量即为聚类数量, 每类的数据求均值作为聚类中心, 根据每组数据和聚类中心的距离可对原始数据进行分类。这种分类方法能够自动确定聚类数量并分类, 无需人工参与, 从而避免人的主观性带来的误差。

将基于膨胀腐蚀的聚类方法运用于风电功率预测中的历史样本 NWP 气象数据分类, 通过对历史样本 NWP 气象数据进行预处理及分类, 根据预测日与每个聚类中心的距离判断预测日所属类别, 采用预测日所属类别的历史样本作为训练样本, 再利用广义回归神经网络(general regression neural

收稿日期: 2016-08-15

基金项目: 国家自然科学基金(51607009)

通信作者: 冬 雷(1967—), 男, 博士、副教授, 主要从事电力电子与电力传动、风力发电与太阳能发电系统, 微电网技术等方面的研究。  
correspondent\_dong@163.com

network, GRNN)对预测日的功率进行预测,并与不分类直接预测功率的预测结果相比较,由仿真结果可看出利用膨胀腐蚀聚类后再预测的精度较高。

## 1 膨胀腐蚀基本理论

### 1.1 相关二值形态学概念

1)二值图:二值图像是指那些灰度只取2个可能值的图像,这2个灰度值通常取为“0”和“1”。习惯上认为取值为“1”的点集 $A$ 对应于图中的物体,而取值为“0”的点集 $A^c$ 构成图的背景<sup>[15]</sup>。

2)结构元素 $B$ :二值形态学的基本思想是利用一个简单预定义的形状对图像进行探测,以得出这个形状是怎样和图形中所包含的形状进行匹配的结论。这种简单的形状就称为结构元素,其本身也是一个二值图像<sup>[16]</sup>。对每个结构元素可以指定一个原点,它是结构元素参与形态学运算的参考点。结构元素在几何结构上比原图像简单且有界。图1给出了3种常用的结构元素,黑色圆点为参考点。

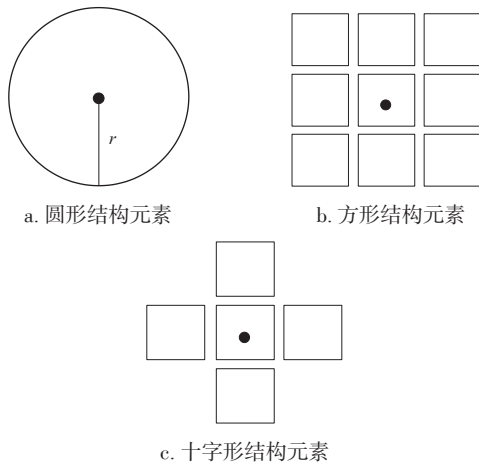


图1 结构元素示意图

Fig. 1 Structuring elements

### 1.2 膨胀和腐蚀

膨胀(dilation)和腐蚀(erosion)变换是建立在集合的 Minkowski 和与差基础上的,是所有复合形态变换或形态分析的基础<sup>[17-19]</sup>。膨胀的直观解释是:将结构元素 $B$ 做映像后,在图像 $A$ 上移动,当 $A$ 与 $B$ 的映像有交集的时候, $B$ 的映像的原点所经过的所有的点构成的集合就是 $B$ 膨胀 $A$ 的结果。腐蚀的直观解释是:当集合 $B$ 完全包含在集合 $A$ 中时, $B$ 的原点位置的集合就是用 $B$ 腐蚀 $A$ 的结果<sup>[15]</sup>。

图2给出膨胀和腐蚀运算的例子,其中图2a中灰色部分是集合 $A$ ;图2b中灰色部分是结构元素 $B$ ;图2c中灰色部分是结构元素 $B$ 的映像;图2d是 $B$ 膨胀 $A$ 后的结果,深色部分为相对原图扩大的部分,可看出此时膨胀运算使图像区域扩张变大;图2e是 $B$ 腐蚀 $A$ 后的结果,深色部分为相对原图剩下的部分,可见腐蚀运算使得图像区域收缩变小。

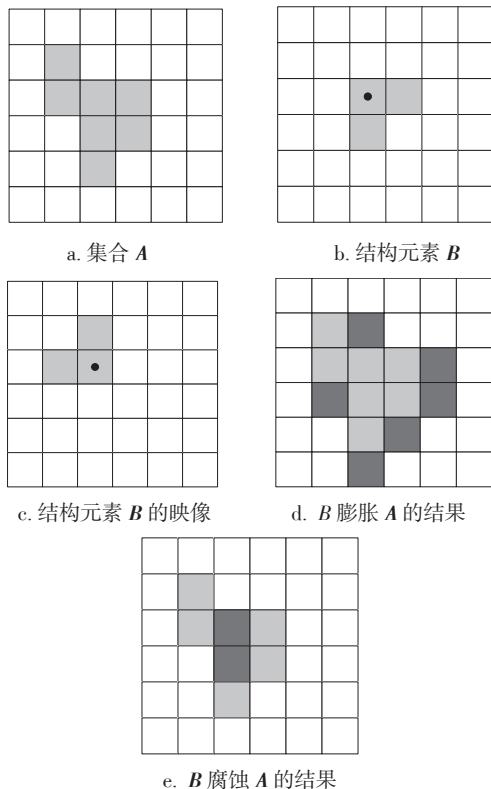


图2 膨胀腐蚀运算示例

Fig. 2 Example of the dilation and erosion

膨胀的作用是填充图像中的小孔及图像边缘处的小凹陷部分,给图像边缘添加像素,使图像增大。腐蚀可把小于结构元素的物体(毛刺、小凸起)去除,这样选取不同大小的结构元素,就可在原图像中去掉不同大小的物体。如果2个物体之间有细小的连通,那么当结构元素足够大时,通过腐蚀运算可将2个物体分开<sup>[20]</sup>。

## 2 基于膨胀腐蚀的聚类方法

聚类的目的是将相似度较大的数据划为一类,使得类内数据相似度最大、差异度最小,类间数据相似度最小、差异度最大。一般认为2组数据之间的距离越大相似度越小,距离越小相似度越大,因

此聚类时应将距离小的 2 组数据划为一类。

在 Matlab 中导入一张二值图,其显示为一个只包含 0 和 1 的二维二值矩阵,因此可变换原始数据集为一个只包含 0 和 1 的二值矩阵,对此矩阵进行膨胀腐蚀,以达到聚类的目的。基于此,本文提出一种新的聚类方法:基于膨胀腐蚀的聚类方法,利用这种聚类方法可自动获取聚类数量。

## 2.1 二维数据集聚类

当待分类数据集是  $n$  组二维数据时,设置一个二维矩阵,处理原始数据使其转换成  $n$  组某一范围内的正整数,那么每组正整数均可对应二维矩阵的一个点;将二维矩阵中的这些点置为 1,其他点置为 0,得到一个只包含 0 和 1 的二维二值矩阵  $A$ ;设置一个二维圆形结构元素  $B$ ,用  $B$  膨胀  $A$  生成新的二值矩阵  $A_1$ ,这样使  $A$  中值为 1 且距离相近的点连成一个区域;设置一个比  $B$  半径大 1 的新的结构元素  $C$ ,再用  $C$  腐蚀  $A_1$ ,这样可去除  $A_1$  中小的区域,同时分开有细小连通的区域,生成新的二值矩阵  $A_2$ ;去掉  $A_2$  中较小的连通区域,聚类数量等于  $A_2$  中剩余的连通区域的数量;提取出每个连通区域对应的整数组并将其归为一类,对每类数据求均值做为这一类的聚类中心;根据与聚类中心的欧式距离对原始数据集分类。

基于膨胀腐蚀确定聚类数量的具体算法如下:

1) 已知数据集有  $n$  组数据点  $(x_i, y_i)$ , 其中  $i=1, 2, 3, \dots, n$ 。按照式(1)将所有数据归一化,生成新的数据点  $(\hat{x}_i, \hat{y}_i)$ :

$$\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \hat{y}_i = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \quad (1)$$

式中,  $x_{\min}$ 、 $x_{\max}$  ——  $x_i$  的最小值和最大值;  $y_{\min}$ 、 $y_{\max}$  ——  $y_i$  的最小值和最大值。

2) 按照式(2)对数据进行处理:首先将  $\hat{x}_i$  扩大  $q_1$  倍,  $\hat{y}_i$  扩大  $q_2$  倍,  $q_j$  是正整数,  $j=1, 2$ , 再取整加 1, 这样所有的数据均可转换为  $1 \sim q+1$  的正整数:

$$x'_i = \text{fix}(\hat{x}_i \times q_1) + 1, y'_i = \text{fix}(\hat{y}_i \times q_2) + 1 \quad (2)$$

式中,  $\text{fix}$  —— 截尾取整函数;  $q_j$  —— 根据原始数据集的范围选择合适的整数。

3) 设矩阵  $A$  为  $(q_1+1) \times (q_2+1)$  的二维零矩阵, 另  $A(x'_i, y'_i) = 1, i=1, 2, 3, \dots, n$ 。这时  $A$  为只包含 0 和 1 的二维二值矩阵。

4) 选择一个半径为  $r$  的圆形结构元素  $B$ ,  $B$  比  $A$  小很多, 用  $B$  膨胀  $A$  得到新的二值矩阵  $A_1$ 。

5) 将  $B$  的半径增加 1 形成新的结构元素  $C$ , 再用  $C$  腐蚀  $A_1$  得到二值矩阵  $A_2$ 。

6) 将矩阵  $A_2$  用二值图显示, 图中较大的连通区域的数量等于聚类数量。

7) 将二值矩阵  $A_2$  中值为 1 且位置相连的点划为一类, 若某类中包含数据数量极少则视为异常情况, 直接舍弃。

8) 取每类所有数据点的平均值作为其聚类中心, 根据与每个聚类中心的欧氏距离确定每组数据所属类别。

## 2.2 三维数据集聚类

当待分类数据集是  $n$  组三维数据时, 设置一个三维矩阵, 处理原始数据使其转换成  $n$  组某一范围内的正整数, 那么每组正整数均可以对应三维矩阵的一个点; 将三维矩阵中的这些点置为 1, 其他点置为 0, 得到一个只包含 0 和 1 的三维二值矩阵  $A$ ; 设置一个三维球形结构元素  $B$ , 先用  $B$  膨胀  $A$  生成新的三维二值矩阵  $A_1$ ; 设置一个比  $B$  半径大 1 的新的结构元素  $C$ , 再用  $C$  腐蚀  $A_1$  生成新的三维二值矩阵  $A_2$ ; 去掉  $A_2$  中较小的连通区域, 聚类数量等于  $A_2$  中剩余的连通区域的数量; 提取出每个连通区域对应的整数组并将其归为一类, 对每类中数据求均值做为这一类的聚类中心; 根据与聚类中心的欧式距离对原始数据集分类。

## 2.3 $m$ 维数据集聚类

当待分类数据集是  $n$  组  $m$  维数据时, 同理可以设置一个  $m$  维矩阵, 每组数据转换为正整数后同样可以对应为  $m$  维矩阵中的一个点, 将  $m$  维零矩阵的这些点的值置为 1, 其他点的值置为 0, 得到一个  $m$  维二值矩阵  $A$ ; 设置一个  $m$  维结构元素  $B$ , 先用  $B$  膨胀  $A$  生成新的  $m$  维二值矩阵  $A_1$ ; 设置一个比  $B$  半径大 1 的新的结构元素  $C$ , 再用  $C$  腐蚀  $A_1$  生成新的  $m$  维二值矩阵  $A_2$ ; 去掉  $A_2$  中较小的连通区域, 聚类数量等于  $A_2$  中剩余的连通区域的数量; 提取出每个连通区域对应的整数组并将其归为一类, 对每类中数据求均值做为这一类的聚类中心; 根据与聚类中心的欧式距离对原始数据集分类。

2.4 UCI数据集的实验

本部分实验采用 UCI 机器学习数据库中的数据集对基于膨胀腐蚀确定聚类数量的算法进行测试。为了验证算法的可行性和普适性,选择属性数量不同的 3 组数据分别进行实验,表 1 是对所用数据集的描述。实验环境是 Matlab 应用软件。

表 1 UCI数据集描述

Table1 Description of UCI datasets

数据集	样本数量/个	属性数量/个	真实聚类数量/个
Far_4k2	400	2	4
Haberman	306	3	2
Iris	150	4	3

2.4.1 Far\_4k2 试验

以数据集 Far\_4k2 为例进行实验:

1)已知数据集 Far\_4k2 有 400 组二维数据点  $(x_i, y_i)$ ,其中  $i=1, 2, 3, \dots, 400$ 。按照式(1)将所有数据归一化。

2)按照式(2)对数据进行操作。因为数据集 Far\_4k2 的 2 个属性值的均值分别为 5.364 和 5.4836, 2 个数值相差很小,所以这里选择  $q_1=q_2=200$ ,这样转换后  $x'_i \in [1,201], y'_i \in [1,201]$ 。

3)设矩阵  $A$  为  $201 \times 201$  的二维零矩阵,另  $A(x'_i, y'_i)=1, i=1, 2, 3, \dots, 400$ 。这时  $A$  为只包含 0 和 1 的二维二值矩阵。

4)选择二维圆形结构元素  $B$  为半径为  $r$  的圆形,取  $r=7$ ,用  $B$  膨胀  $A$  得到  $A_1$ ,将  $r$  加 1 得到半径为 8 的圆形结构元素  $C$ ,再用  $C$  腐蚀  $A_1$  得到  $A_2$ 。在 Matlab 中用图像显示二值矩阵  $A_2$  结果如图 3 所示,可看出此数据集可以分为 4 类。



图 3 Far\_4k2数据集的实验结果  
Fig. 3 Simulation result of Far\_4k2

2.4.2 Haberman 和 Iris 试验

由于各种数据集的样本数量和空间分布不同,因此参数的选择也不同。数据集 Haberman 和 Iris 的参数如表 2 所示。其中 Haberman 数据集的 3 种属性值的均值分别为 52.45、62.85 和 4.03,因此选择  $q_1=q_2=50, q_3=30$ ;Iris 数据集是四维数据,为了减少计算量,其  $q_i$  值不宜设置很大。Haberman 数据集利用膨胀腐蚀的方法确定聚类数量的结果如图 4 所示。

表 2 数据集实验参数

Table 2 Description of UCI datasets

数据集	$q_i$	$r$
Haberman	$q_1=q_2=50, q_3=30$	6
Iris	$q_1=30, q_2=25, q_3=20, q_4=30$	3

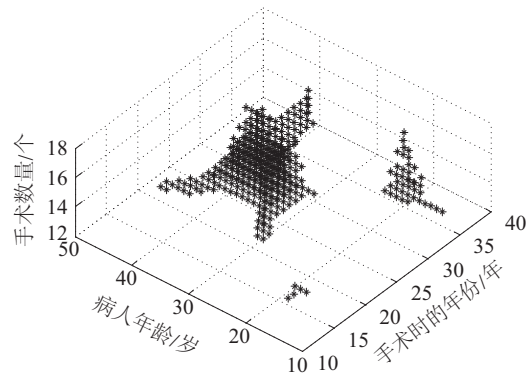


图 4 Haberman的实验结果  
Fig. 4 Simulation results of Haberman

由于数据集 Iris 的属性数量为 4,因此当膨胀腐蚀完后的矩阵为 4 维矩阵,在 Matlab 中无法通过图形显示,提取出矩阵中相连的值为 1 的点划为一类,其结果如表 3 所示,可看出 Iris 数据集可分为 3 类。

表 3 Iris提取相邻数据的结果

Table 3 Iris result of extracting neighboring data

类别	第 1 类	第 2 类	第 3 类
包含数据组	4739	19410	759

3 基于膨胀腐蚀聚类的风电功率预测

3.1 预测模型

将膨胀腐蚀的聚类方法用于风电功率预测中的历史样本 NWP 气象数据聚类,选择与预测日同一类的样本数据作为的训练样本,最后利用广义回



归神经网络建立预测模型。预测模型如图 5 所示。对模型进行训练时,采用预测日所属分类样本的 NWP 气压、NWP 风速、NWP 气温、NWP 风向余弦、NWP 风向正弦作为输入,功率作为输出。将预测日的 NWP 信息输入到以训练好的神经网络中得到风力发电功率的预测值。其中预测尺度为 1 d,预测步长为 96。

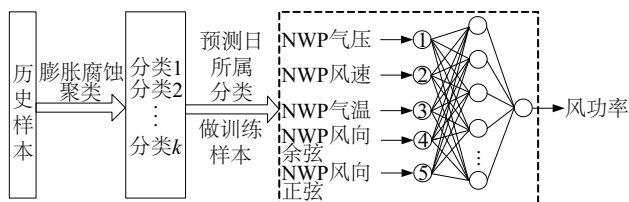


图5 预测模型

Fig. 5 Prediction model

### 3.2 实验仿真及分析

仿真数据描述:测试数据集来源于黑龙江依兰风电场,数据包括风速、气温、风向、气压、功率 5 个属性值,时间长度为 2012 年 1~6 月份共 182 d,时间分辨率为 15 min。对数据进行处理后得到每天

的气压均值、最小风速、最大风速、最低气温、最高气温、风向正弦均值、风向余弦均值、风速均值 8 个属性值代表天气信息。

1)利用基于膨胀腐蚀的方法对历史样本 NWP 气象数据进行分类:由于在风电预测中气象数据分类的最终目的是更精确的预测发电功率,因此为了减少计算量首先按照式(3)分别计算气象数据集中 8 个属性与功率的相关度,结果如表 4 所示,选取相关度较大的最小风速、最大风速和风速均值 3 个属性作为膨胀腐蚀聚类方法的输入对历史样本 NWP 气象数据进行分类;按照前文所述的膨胀腐蚀聚类数量确定方法对历史气象数据进行分类,其中选取参数  $q_1 = 40$ ,  $q_2 = 60$ ,  $q_3 = 50$ , 半径  $r = 4$ 。

$$r_{x,y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad (3)$$

聚类结果如图 6 所示,历史 181 d NWP 样本气象数据分为 3 类,其中第 1 类包括 78 d,第 2 类包括 38 d,第 3 类包括 65 d。

表4 8种气象属性与功率的相关度

Table 4 Correlation degree between 8 meteorological factors and power

气象属性	气压均值	最小风速	最大风速	最低气温	最高气温	风向正弦均值	风向余弦均值	风速均值
与功率的相关度	0.107	0.715	0.807	-0.399	-0.377	-0.011	-0.049	0.938

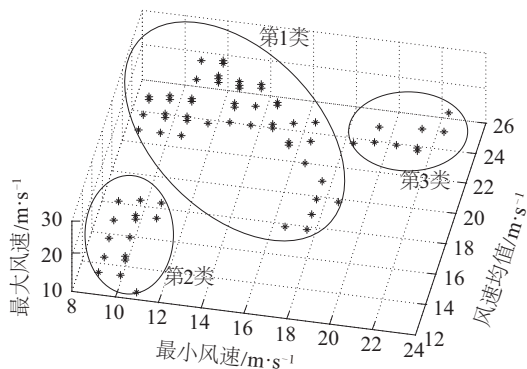


图6 风电预测的历史天分类

Fig. 6 Clustering result of historical days

2)确定预测日所属类别。选择第 182 天作为预测日,根据第 182 天与 3 个聚类中心的距离(分别为 $[7.9981, 7.5988, 9.4471]$ ),由此可判断出预测日属于第 2 类,在第 2 类样本中选择离预测日时间最近的 15 个历史天作为预测样本。

3)对预测日进行预测。将气压、风速、气温、风

向余弦、风向正弦作为 GRNN 神经网络的输入,功率数据作为输出,利用 Matlab 进行仿真得到预测日的功率预测结果。

4)不对历史气象数据进行分类,直接应用 GRNN 模型进行预测,预测结果曲线对比如图 7 所示,误差对比如图 8 所示。

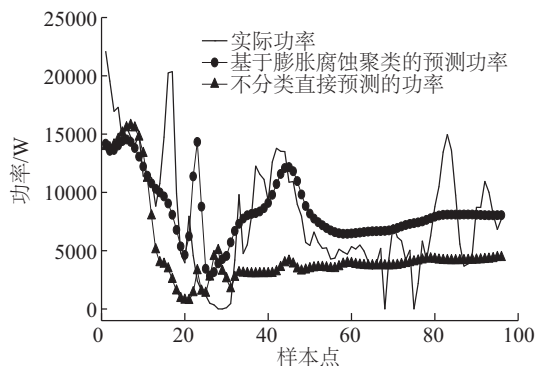


图7 预测结果曲线对比

Fig. 7 Comparison of wind power prediction

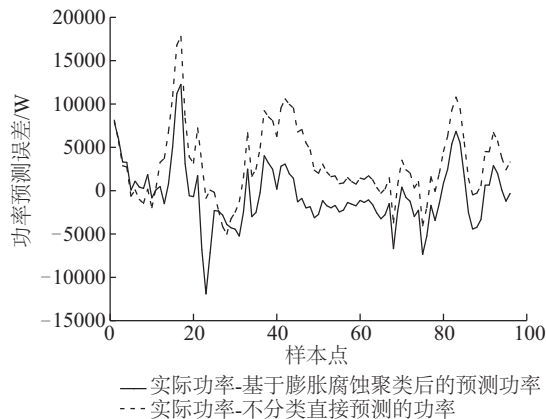


图8 预测误差曲线对比

Fig. 8 Comparison of two kinds of prediction errors

5) 根据式(4)和式(5)计算标准绝对值平均误差  $NMAE$  和标准均方根误差  $NRMSE$ , 对预测结果进行评估, 如表5所示。

表5 预测误差的统计指标

Table 5 Statistical indices of prediction errors

误差	$NMAE/\%$	$NRMSE/\%$
分类预测	5.9929	7.9371
不分类直接预测	9.5265	12.9580

$$NMAE = \frac{1}{n} \sum_{i=1}^n \frac{|p_i - p'_i|}{P_{cap}} \times 100\% \quad (4)$$

$$NRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{p_i - p'_i}{P_{cap}} \right)^2} \quad (5)$$

式中,  $n$  —— 预测样本数量;  $p_i$ 、 $p'_i$  —— 样本实际值和预测值;  $P_{cap}$  —— 风电场的额定装机容量。

由图7和表4可看出利用膨胀腐蚀的聚类方法对历史样本 NWP 气象数据进行分类后, 选取与预测日同一类的历史天作为训练样本得到的预测模型, 其预测精度要高于不分类直接进行训练得到的预测模型。

## 4 结 论

风电功率变化趋势近似日的气象变化趋势也近似, 因此对历史 NWP 气象样本数据进行分类后, 提取与预测日 NWP 信息相似性高的历史样本作为预测模型的训练输入, 可在提高风电功率预测精度的同时大大减小建模仿真的计算量。本文提出基于膨胀腐蚀的聚类方法, 通过对 UCI 数据集进行分类, 验证此方法的可行性。将其应用于

风电功率预测中的历史样本 NWP 气象数据分类, 利用膨胀腐蚀的聚类方法对历史天进行分类后, 选取与预测日同一类的历史天作为训练样本得到的预测模型, 其预测精度高于不分类直接进行训练得到的预测模型。又因为此种聚类方法无需人工参与, 因此在风电功率预测时可保证聚类的实时性。

## [参考文献]

- [1] Dong Lei, Wang Lijie, Khahro S F, et al. Wind power day-ahead prediction with cluster analysis of NWP[J]. Renewable and Sustainable Energy Reviews, 2016, 60: 1206—1212.
- [2] 周 涛, 陆惠玲. 数据挖掘中聚类算法研究进展[J]. 计算机工程与应用, 2012, 48(12): 100—111.
- [2] Zhou Tao, Lu Huiling. Clustering algorithm research advances on data mining[J]. Computer Engineering and Application, 2012, 48(12): 100—111.
- [3] 关晓惠, 钱亚冠, 孙欣欣. 一种改进的基于局部密度的聚类算法[J]. 电信科学, 2016, 32(1): 54—59.
- [3] Guan Xiaohui, Qian Yaguan, Sun Xinxin. An improved clustering algorithm based on local density[J]. Telecommunications Science, 2016, 32(1): 54—59.
- [4] 孙吉贵, 刘 杰, 赵连宇. 聚类算法研究[J]. 软件学报, 2008, 19(1): 48—61.
- [4] Sun Jigui, Liu Jie, Zhao Lianyu. Clustering algorithms research[J]. Journal of Software, 2008, 19(1): 48—61.
- [5] 周世兵, 徐振源, 唐旭清. 新的 K-均值算法最佳聚类数确定方法[J]. 计算机工程与应用, 2010, 46(16): 27—31.
- [5] Zhou Shibing, Xu Zhenyuan, Tang Xuqing. New method for determining optimal number of clusters in K-means clustering algorithm[J]. Computer Engineering and Application, 2010, 46(16): 27—31.
- [6] 韩凌波. K-均值算法中聚类个数优化问题研究[J]. 四川理工学院学报: 自然科学版, 2012, 25(2): 77—80.
- [6] Han Lingbo. Optimization study on class number of K-means Algorithm[J]. Journal of Sichuan University of Science & Engineering: Natural Science Edition, 2012, 25(2): 77—80.
- [7] 魏建东, 陆建峰, 彭甫谔. 一种层次初始的聚类个数自适应的聚类方法研究[J]. 电子设计工程, 2015, 23(6): 5—8.

- [7] Wei Jiandong, Lu Jianfeng, Peng Furong. Research on a method of self-adaptation of the number of clusters for hierarchical initialization clustering[J]. Electronic Design Engineering, 2015, 23(6): 5—8.
- [8] Řezanková H, Húsek D. Fuzzy clustering: Determining the number of clusters [A]. Computational Aspects of Social Networks, 2012 Fourth International Conference on IEEE[C], Sao Carlos, 2012, 277—282.
- [9] 孙昌思核, 孔万增, 戴国骏. 一种自动确定类个数的谱聚类算法[J]. 杭州电子科技大学学报, 2010, 30(2): 53—56.
- [9] Sunchang Sihe, Kong Wanzeng, Dai Guojun. A spectral clustering with ascertainable clustering number [J]. Journal of Hangzhou Dianzi University, 2010, 30(2): 53—56.
- [10] Zhang Yangfeng, Xu Xiaofei, Ye Yunming. NSS-A  $K$ -means: An agglomerative fuzzy  $K$ -means clustering method with automatic selection of cluster number [A]. Advanced Computer Control (ICACC), 2010 2nd International Conference on IEEE[C], Shenyang, 2010, 32—38.
- [11] 谢娟英, 屈亚楠. 密度峰值优化初始中心的  $K$ -medoids 聚类算法[J]. 计算机科学与探索, 2016, 10(2): 230—247.
- [11] Xie Juanying, Qu Yanan.  $K$ -medoids clustering algorithms with optimized initial seeds by density peaks [J]. Journal of Frontiers of Computer Science and Technology, 2016, 10(2): 230—247.
- [12] Muneeswaran P, Velvizhy P, Kannan A. Clustering fusion with automatic cluster number [A]. Recent Trends in Information Technology (ICRTIT), 2014 International Conference on IEEE[C], Chennai, 2014, 1—6.
- [13] Wang Cangdong, Lai Jianhuang, Huang Dong. Kernel-based clustering with automatic cluster number selection [A]. 2011 IEEE 11th International Conference on Data Mining Workshops[C], Vancouver, 2011, 293—299.
- [14] Gao Mingming, Chang Taihua, Gao Xiangxiang. Research in data stream clustering based on Gaussian mixture model genetic algorithm [A]. The 2nd International Conference on Information Science and Engineering[C], Wuhan, 2010, 3904—3907.
- [15] 文 华. 基于数学形态学的图像处理算法的研究 [D]. 哈尔滨: 哈尔滨工程大学, 2007.
- [15] Wen Hua. Research of image processing algorithm based on mathematic morphology [D]. Harbin: Harbin Engineering University, 2007.
- [16] 孙兴华, 郭 丽. 数字图像处理——编程框架, 理论分析, 实例应用和源码实现 [M]. 北京: 机械工业出版社, 2012.
- [16] Sun Xinghua, Guo Li. Digital image processing—Programming framework, theoretical analysis, examples of applications and source code [M]. Beijing: China Machine Press, 2012.
- [17] 阮秋琦. 数字图像处理学 [M]. 北京: 电子工业出版社, 2001.
- [17] Ruan Qiuqi. Digital image processing [M]. Beijing: Publishing House of Electronics Industry, 2001.
- [18] Guy R. Cochrane, Kevin D. Lafferty. Use of acoustic classification of sidescan sonar data for mapping benthic habitat in the Northern Channel Islands [M]. California: Continental Shelf Research, 2000.
- [19] Agam G. Regulated morphological operations [J]. Pattern Recognition, 1999, 32: 133—134.
- [20] 何东健, 耿 楠, 张 义. 宽数字图像处理 [M]. 西安: 西安电子科技大学出版社, 2003.
- [20] He Dongjian, Geng Nan, Zhang Yi. Digital image preprocess [M]. Xi'an: Xi'an Electronic Science and Technology University Press, 2003.

## WIND POWER PREDICTION BASED ON DILATION AND EROSION CLUSTERING METHOD

Zhou Xiao<sup>1</sup>, Dong Lei<sup>1</sup>, Hao Ying<sup>1</sup>, Liao Xiaozhong<sup>1</sup>, Gao Yang<sup>2</sup>

(1. *College of Automation, Beijing Institute of Technology, Beijing 100081, China;*

2. *College of Electric Power, Shenyang Institute of Engineering, Shenyang 110136, China*)

**Abstract:** A new clustering method based on dilation and erosion is proposed and UCI (University of California Irvine) data set is used to carry out experimental simulation to prove the feasibility of this method. Then this clustering method is used to classify NWP (numerical weather prediction) information in wind power prediction, selecting the historical day data of the same type as the forecast day data as the training sample, and the generalized regression neural network is used to predict the power and compare with the direct prediction method. The simulation results show that the re-prediction has higher prediction accuracy after classification of historical day data based on dilation and erosion clustering analysis.

**Keywords:** wind power prediction; clustering; dilation and erosion; generalized regression neural network (GRNN)