

The Study of Short-Term Wind Power Generation Prediction Methods Based on Clustering Analysis in Wind Farm

Cun Dong¹, Shuang Gao², Ying Hao², Yang Gao³

¹State Grid Corporation of China, Beijing

²Beijing Institute of Technology, Beijing

³Shenyang Institute of Engineering, Shenyang Liaoning

Email: Correspondent_dong@163.com

Received: Jun. 2nd, 2015; accepted: Jun. 20th, 2015; published: Jun. 23rd, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In order to make full use of information contained from historical wind data, this paper proposed a short-term power prediction method based on clustering analysis according to the daily similarity property of wind speed and wind power. By the preprocessing of the original sample data, when calculating the Euclidean distance among the characteristic parameters, the history data which are similar to the NWP characteristic parameter of the prediction day are used as the training sample; the NWP information provided by the meteorological department is used as the characteristic parameter of the prediction day. This Euclidean distance is used to be the basis of the similarity measure. Finally, this paper uses the similar samples after clustering to establish the short-term prediction model which is using the NWP data as the input parameter, using the actual wind power generation as the target data. Testing by the actual wind farms, the prediction precision is improved obviously.

Keywords

Short-Term Wind Power Generation Prediction, Clustering Analysis, K-Means Clustering Method, Daily Similarity Property, Numerical Weather Prediction (NWP)

基于聚类分析的风电场短期功率预测方法研究

董存¹, 高爽², 郝颖², 高阳³

¹国家电网公司, 北京

²北京理工大学, 北京

³沈阳工程学院, 辽宁 沈阳

Email: Correspondent_dong@163.com

收稿日期: 2015年6月2日; 录用日期: 2015年6月20日; 发布日期: 2015年6月23日

摘要

为了充分利用历史风速数据所蕴含的信息, 本文根据风速和风电功率的日相似性提出基于聚类分析的短期功率预测方法, 通过对原样本数据进行预处理, 选取与预测日NWP特征参数相似的历史日数据, 以此作为建立模型的训练样本, 将气象部门提供预测日的NWP信息作为预测日的特征参数, 计算特征参数间的欧式距离, 以此作为相似性度量的依据, 最后利用聚类后的相似样本建立预测模型, 以NWP数据为输入参数, 实际风电功率为目标值, 经过训练后得到聚类风电功率短期预测模型。经实际风电场测试, 预测精度明显提高。

关键词

短期功率预测, 聚类分析, K均值聚类法, 日相似性, 数值天气预报(NWP)

1. 引言

风速具有明显特征区别, 同时每天的风速随着日出日落的周期变化亦存在一定的规律, 经大量研究发现, 风速存在一定的日相似特性[1]。为了充分利用历史风速数据所蕴含的信息, 根据风速和风电功率的日相似特性提出基于聚类分析的短期风电功率预测方法, 出发点在于利用样本预处理对改善预测结果的重要作用, 最终目的在于提高预测精度。

采用聚类方法进行风电功率预测的关键是寻找聚类的依据, 即获得不同类型样本划分的依据。目前, 国内外针对风电功率短期预测及其预测精度提高方法的研究亦较多, 但是采用相似日聚类方法作为预测前处理, 以提高预测精度的方法还未见相关研究。风电功率变化趋势近似的日, 存在某些共有的特征, 可通过 NWP 信息变化趋势的基本一致性得到体现。

本文所提方法将聚类分析法应用于风电功率预测, 通过对原样本数据进行预处理, 提取与预测日 NWP 特征参数相似的历史日数据作为建模的训练样本。以预测日的 NWP 信息作为预测日特征参数, 通过计算所获得特征参数的欧氏距离确定各类相似日数据, 最后利用聚类后的相似样本建立功率预测模型, NWP 数据作为模型输入参数, 实际风电功率数据作为模型目标值, 经过模型训练便能得到风电场短期功率多步预测模型。

2. 风电功率的日相似性

风电场风电机组实际输出功率的大小主要取决于当地风能资源情况, 而风资源的特性主要是指风速的变化特性。风速是单位时间内空气在水平方向上移动的距离, 主要受气象因素及地形、地表障碍物等因素的影响。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, 所以天气状况相似日的风电功率变化趋势十分近似。

图 1 是依兰风电场 2012 年 1 月至 6 月, 部分日风速变化曲线族, 图中横坐标表示风速采集时间点,

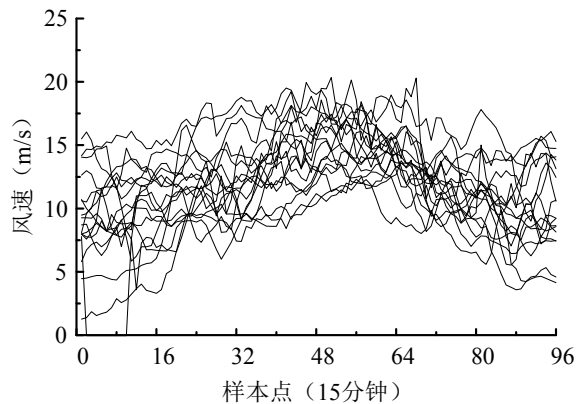


Figure 1. Some of the daily wind speed change curve
图 1. 部分日风速变化曲线族

采集间隔为 15 min，可以看出，其变化趋势存在一定的相似性，对应的风电功率亦呈现相似的变化趋势，如图 2 所示，图中横坐标表示与图 1 相对应的风电功率采集时间点，采集间隔同样为 15 min。由于风电功率的变化趋势与该日的天气状况存在一定关系，对于某风电场来说，风电功率的日相似性可采用与其有密切关系的天气信息的日相似情况来判断。气象信息可通过预报的方式获得，根据其相似性查找预测时段的相近样本，作为建立短期预测模型的训练样本。聚类分析法是解决相似性样本查找的一种有效方法。

3. 相似日的聚类分析

聚类分析是将研究对象按照一定度量标准分成不同类别的统计分析技术[2]。聚类是属于无监督的分类方法，是一种探索性的分析，没有预先指定的类别。在机器学习领域，聚类是无指导学习。聚类分析被广泛应用于多个领域，包括模式识别、数据分析、图像处理、以及市场研究。在商业、网络以及生物应用等方面都发挥着重要的作用。

3.1. 聚类分析的基本步骤

运用聚类方法一般包括下列四个步骤[3]-[5]：

1) 找出分类对象的特征；

一般，所研究的数据对象包含多种特征，其中具有代表性的特征能够帮助识别数据对象本身，这种特征属性是区分不同数据对象的标准，所以在进行聚类分析时，首先需将对象区别于其他数据的特点提取出来。

2) 确定分类对象的相似性；

聚类就是把特征属性相类似的数据对象划分到一类，特征属性不同的划分到另一类。所以衡量数据的相似性要有一个度量的标准。不同的标准得出的聚类结果可能是不同的。关于相似性的量度在 2.3 节中有介绍。

3) 提出达到分类的步骤或聚类算法；

聚类过程中的关键步骤之一是采用不同的方法对数据对象进行划分。主要的聚类分析方法有划分方法和层次方法。而硬聚类和模糊聚类是划分方法中比较常用的两种技术。

4) 对聚类结果做出评价。

评价聚类结果质量的好坏是没有客观标准的。可以根据实际情况选择更符合实际的聚类结果。

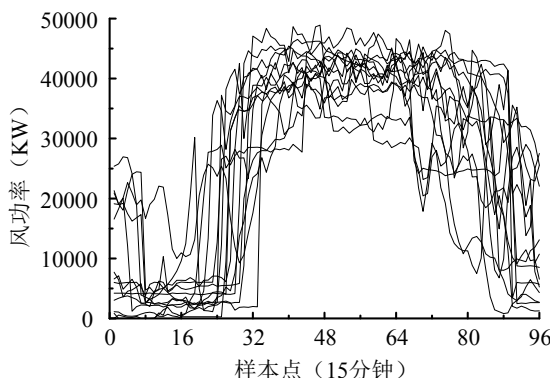


Figure 2. Some of the daily trend curve of the wind power generation
图 2. 风电功率部分日功率趋势变化曲线族

3.2. 聚类分析的方法

传统的聚类算法包括五类：划分方法、层次方法、基于密度的方法、基于网格的方法和基于模型的方法[6]-[8]。划分方法和层次方法都是基于距离来判断相似度的，基于密度的方法是基于密度这一概念来判断相似度的。现代的聚类算法包括高维聚类分析方法和动态聚类算法两类。

K 均值聚类法是动态聚类算法中最经典的一种，其基本思想是将每一个样本划分到离均值最近的类别中，它是以距离的远近为标准进行聚类的。

K 均值聚类算法一般包括以下几个处理步骤：

- 1) 将所有数据分为 K 个初始类，选取 K 个样本点为初始聚类中心，记为 $z_1(l), z_2(l), \dots, z_k(l)$ ，其中初始值 $l=1$ ；
- 2) 按照最近邻规则将所有样本分配到各聚类中心所代表的 K 类 $\omega_j(K)$ 中，各类所包含的样本数为 $N_j(l)$ ；
- 3) 计算各类的均值向量，并将该向量作为新的聚类中心：

$$z_j(l+1) = \frac{1}{N_j(l)} \sum_{x(i) \in \omega_j(l)} x(i) \quad (1)$$

其中 $j=1, 2, \dots, k$ ， $i=1, 2, \dots, N_j(l)$ 。

- 4) 若 $z_j(l+1) \neq z_j(l)$ ，表示聚类结果并不是最佳的，则返回步骤 2)，继续迭代计算；
- 5) 若 $z_j(l+1) = z_j(l)$ ，迭代过程结束，此时的聚类结果就是最优聚类结果。

3.3. 相似性的量度

在开始对样本进行分组之前，必须定义相似性的量度。每一个样本必须与其他样本相比较，然后才能将非常“相似”的样本置于同一类中，不相似的样本则置于不同的类中。两个元素之间的相似性可用各种不同的方法来度量，主要有距离度量、相关性度量和信息量度量等方法。

常用的距离有明氏距离(Minkowski)、马氏距离(Mahalanobis)和兰式距离(Canberra)。

本文采用欧式距离度量不同日的风速趋势之间的相似程度。将每天作为一个数据对象，由一个 7 维向量表示，称为日 NWP 向量，表示为 $X = [Pav, Vmin, Vmax, Tmin, Tmax, Dsin, Dcos]$ ，其中的变量依次代表日气压平均值、日风速最小值、日风速最大值、日气温最小值、日气温最大值、日风向正弦平均值、日风向余弦平均值。由于日 NWP 向量中各分量的量纲不同，需要进行归一化处理，气压、风速和气温

分别除以各自的历史最大值，风向正弦和余弦值均为归一化数值，不需再作处理。

距离定义为：

$$d_i = \left(\sum_{k=1}^7 (x_m(k) - x_i(k))^2 \right)^{1/2} \quad (2)$$

其中， d_i 为预测日与历史样本 i 的欧氏距离。 x_m 为预测日的日 NWP 向量， x_i 为历史数据的日 NWP 向量， $i = 1, 2, \dots, n$ ， n 为样本数。

4. 基于聚类分析的短期风电功率预测模型

4.1. 模型结构

图 3 为基于聚类分析的短期风电功率预测模型的结构示意图。首先利用 K 均值聚类算法以日 NWP 向量距离最接近原则将历史样本分为 K 类，采用计算机编程自动搜寻预测日所属的分类，将各分类中的样本数据作为训练样本，利用神经网络建立预测模型。

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 风向正弦和 NWP 风向余弦等作为输入，模型的目标值为风电场输出功率。采用此模型进行预测时，只需将预测日的 NWP 信息输入模型便可得到风电功率的预测值。

4.2. 聚类计算与结果分析

采用黑龙江依兰风电场 2012 年 1~2 月的 NWP 数据和实测风电功率数据进行分析、建模和预测。数据分辨率为 15 分钟。选择 2012 年 2 月 4 日作为预测日，预测步长为 96 点。

选择 2012 年 2 月 4 日之前的 20 天历史数据做聚类分析，采用 K 均值聚类算法，得到准则函数与分类数 K 的关系曲线如图 4 所示，根据 2.2 节中介绍的确定最优分类数的方法，取准则函数曲线拐点处的 K 作为最佳分类数，得到 $K = 3$ 。

在分类数 $K = 3$ 的情况下，20 个历史样本日所属类别情况如表 1 所示。其中，4 天属于第 3 类，1 天属于第 2 类，其他均属于第 1 类。由式(1)可计算出这三类的聚类中心(归一化)分别为：

第 1 类：[0.988 0.183 0.438 -1.130 -0.804 0.042 0.051]；

第 2 类：[0.988 0.555 0.863 -1.151 -0.853 0.119 0.189]；

第 3 类：[0.9930.047 0.268 -0.856 -0.551 -0.020 -0.125]。

预测当日和 2012 年 2 月 4 日的归一化日 NWP 向量为[0.981 0.340 0.801 -0.932 -0.579 0.113 -0.052]，由式(2)可得与三类聚类中心的欧式距离分别为 0.51、0.48 和 0.63，距离第 2 类聚类中心最近，所以预测日所属分类为第 2 类，属于第 2 类的样本为第 19 个样本日，2012 年 2 月 2 日。

4.3. 聚类结果用于短期预测

以 2012 年 2 月 2 日 NWP 的气压、风速、气温、风向正弦和风向余弦作为输入，实测功率数据作为输出进行建模，数据分辨率为 15 分钟，训练样本数据个数为 96，神经网络结构采用 GRNN，窗口宽度参数 s 取值为 0.15，模型结构见图 3。

模型训练完成后，将预测日即 2012 年 2 月 4 日的 NWP 气压、NWP 风速、NWP 气温、NWP 风向正弦和 NWP 风向余弦输入得到预测功率值。预测误差 NMAE 和 NRMSE 分别为 10.67%和 14.01%。

为了验证本文方法的先进性，将基于聚类分析的神经网络模型与持续模型进行对比，得到图 5 所示预测结果对比曲线。功率预测误差情况详见表 2。

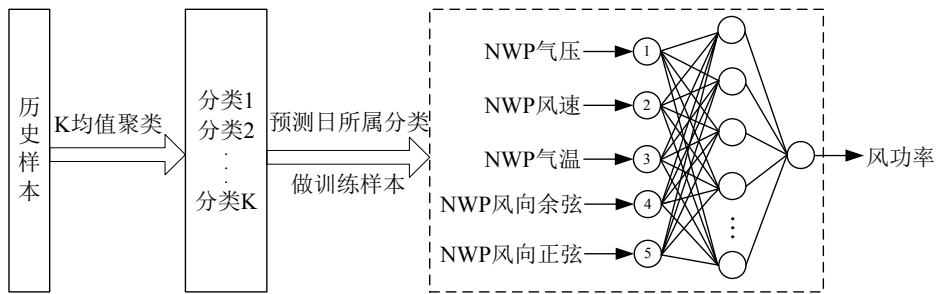


Figure 3. Short-term wind power generation prediction mode based on clustering analysis
图 3. 基于聚类分析的短期风电功率预测模型

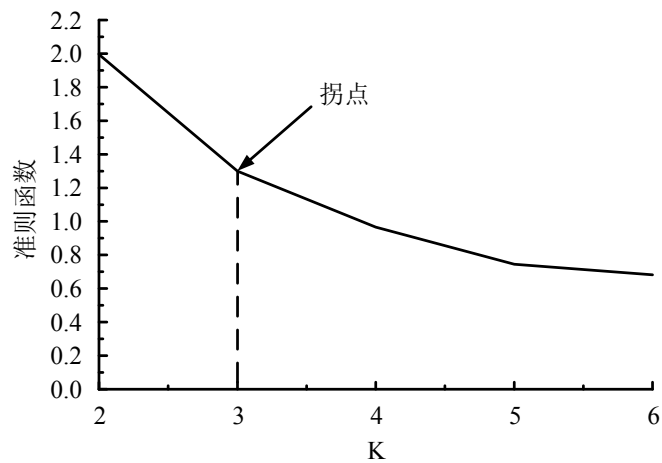


Figure 4. The diagram of the criterion function and the classification number K
图 4. 准则函数与分类数 K 关系图

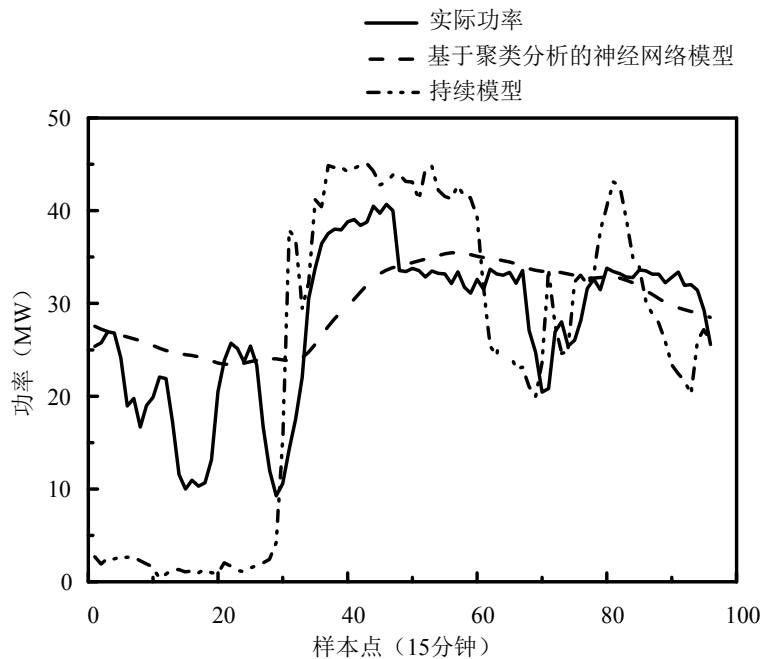


Figure 5. The comparison of the prediction curve from different prediction mode
图 5. 各模型功率预测曲线对比

Table 1. The sample clustering table
表 1. 样本所在聚类情况表

样本日	1	2	3	4	5	6	7	8	9	10
所属类别	1	3	1	3	3	3	1	1	1	1
样本日	11	12	13	14	15	16	17	18	19	20
所属类别	1	1	1	1	1	1	1	1	2	1

Table 2. The comparison of the prediction error from different prediction mode
表 2. 功率预测误差

风电功率预测模型	NMAE/%	NRMSE/%
基于聚类分析的神经网络模型	10.67	14.01
持续模型	22.12	26.81

表 2 中标准绝对值平均误差(normalized mean absolute error, NMAE), 标准均方根误差(normalized root mean square error, NRMSE)计算公式如式(3)、(4):

$$NMAE = \frac{1}{P_{cap}} \times \frac{1}{N} \sum_{i=1}^N |x'(i) - x(i)| \times 100\% \quad (3)$$

$$NRMSE = \frac{1}{P_{cap}} \times \sqrt{\frac{1}{N} \sum_{i=1}^N (x'(i) - x(i))^2} \times 100\% \quad (4)$$

上述三式中 $x(i)$ 为样本真实值, $x'(i)$ 为样本预测值, P_{cap} 是风电场的额定装机容量, N 为预测样本个数。

5. 结论

由于一年四季的风速特征具有明显区别, 且每天的风速随着日出日落的周期变化也存在一定的规律, 为了充分利用历史风速数据所蕴含的信息, 本文分析了风电功率的日相似性, 而对应的气象数据也具有相似性, 气象数据是可以提前获取预报值的, 这就为利用相似性这一特征进行风电功率短期预测提供了思路。经过实际风电场测试, 验证了本文所提方法的有效性和先进性, 具体结论为:

- 1) 对风电功率预测模型进行恰当的前处理能够有效的提高短期风电功率的预测精度;
- 2) 采用聚类分析方法对 NWP 数据进行分析可有效的对不同天气类型的日数据进行分类;
- 3) 将聚类分析方法与神经网络进行衔接可建立有效的短期风电功率预测模型, 提高预测精度。

参考文献 (References)

- [1] 朱峰, 孙辉, 周玮 (2008) 基于相似日聚类的神经网络风速预测. *新能源发电, 中国高等学校电力系统及其自动化专业第二十四届学术年会论文集*, 北京, 2008 年 10 月 24 日, 2627-2630.
- [2] 殷君伟 (2013) K-均值聚类算法改进及在服装生产的应用研究. 硕士学位论文, 苏州大学, 苏州.
- [3] 李尤丰 (2009) 图像处理技术在硬度测量系统中的应用研究. 硕士学位论文, 南京理工大学, 南京.
- [4] 黄山山 (2011) 南京电信网厅用户行为分析系统的设计实现与应用. 硕士学位论文, 南京理工大学, 南京.
- [5] 古家声 (2012) 数据挖掘技术在寿险客户分析中的应用. 硕士学位论文, 华南理工大学, 广州.
- [6] 刘蕴 (2009) 基于启发式方法的集成生产-配送调度研究. 硕士学位论文, 天津大学, 天津.
- [7] 骆永健 (2010) 基于聚类的数据流异常检测算法的研究. 硕士学位论文, 哈尔滨工程大学, 哈尔滨.
- [8] 汪瑛 (2009) 数据挖掘在燃气系统中的应用研究. 硕士学位论文, 南京理工大学, 南京.