

# 基于 K-means 和改进 KNN 算法的 风电功率短期预测系统

何建强<sup>1,2</sup>, 张玉萍<sup>3</sup>, 滕志军<sup>4</sup>

(1. 商洛学院, 陕西 商洛 726000; 2. 商洛市人工智能研究中心, 陕西 商洛 726000;

3. 北京四方继保自动化股份有限公司, 北京 100085;

4. 东北电力大学 现代电力系统仿真控制与绿色电能新技术教育部重点实验室, 吉林 吉林 132012)

**摘要:** 为提高风电功率短期预测的准确性, 针对 KNN (K-Nearest neighbor algorithm) 算法在风电功率预测中的不足, 提出了基于 K-means 和改进 KNN 算法的风电功率短期预测方法; 利用 K-means 聚类方法确定风电历史样本的类别, 对 KNN 算法中搜索相似历史样本集的方式进行了改进和优化, 构建了预测模型, 并采用 C/S 架构实现了预测系统的设计; 该系统具有自修正功能, 能够随着预测次数的增加, 不断修正预测模型, 逐渐降低预测的误差率; 以吉林省某风电场历史数据为样本进行了仿真分析, 结果显示该算法与其它算法相比平均绝对误差和均方根误差最大下降 1.08% 和 0.48%, 运算时间提升了 5.45%, 在风电功率超短期多步预测中具有推广应用价值。

**关键词:** 风电功率; 预测; K-means 聚类算法; 改进 KNN 算法

## Wind Power Short-term Forecasting System Based on K-means and Improved KNN Algorithm

HE Jianqiang<sup>1,2</sup>, ZHANG Yuping<sup>3</sup>, TENG Zhijun<sup>4</sup>

(1. Shangluo University, Shangluo 726000, China; 2. Shangluo Artificial Intelligence Research Center, Shangluo 726000, China; 3. Beijing Sifang Automation Co., Ltd., Beijing 100085, China; 4. Northeast Electric Power University, Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Jilin 132012, China)

**Abstract:** In order to improve the accuracy of short-term prediction of wind power, in view of the shortcomings of KNN (K-Nearest Neighbor algorithm) algorithm in wind power prediction, based on K-means and an improved KNN algorithm, a short-term wind power forecasting method is proposed. The K-means clustering method is used to determine the types of historical wind power samples, the method of searching for similar historical sample sets in the KNN algorithm is improved and optimized, a prediction model is constructed, and the C/S architecture is used to realize the design of prediction system. The system has a self-correction function, which can continuously correct the forecast model as the number of forecasts increases, and gradually reduce the error rate of forecast. A simulation analysis with historical data of a wind farm in Jilin Province is carried out. The results show that, compared with other algorithms, the algorithm has the largest decrease in average absolute error and root mean square error by 1.08% and 0.48%, and the calculation time has increased by 5.45%, the ultra-short-term multi-step forecasting has the value of promotion and application.

**Keywords:** wind power; prediction; K-means clustering algorithm; improved KNN algorithm

## 0 引言

风能作为一种可再利用率、清洁的新型替代能源, 能够优化能源结构, 增强能源供应能力, 保障能源安全可靠。由于风具有随机性、间歇性和不可控性, 导致风电具有间

断性, 致使风电并网会增加电力系统运行的不稳定性, 极大地影响电网的供电质量。因此, 科学、准确的预测风电场的输出功率能够平衡大规模风电对电网的影响<sup>[1]</sup>, 促进风电健康持续地发展。文献 [2] 和文献 [3] 阐述了由于风电的间断性, 致使风电功率预测大多主要集中于短期预

收稿日期: 2021-11-03; 修回日期: 2021-12-10。

基金项目: 国家自然科学基金项目(61501107); 陕西省教育厅 2019 年度专项科学研究计划项目(19JK0261); 商洛学院服务地方科研专项项目(19FK002)。

作者简介: 何建强(1987-), 男, 陕西商州人, 硕士, 讲师, 主要从事新能源发电功率预测及优化控制方向的研究。

滕志军(1973-), 男, 吉林省吉林市人, 教授, 博导, 主要从事无线通信技术方向的研究。

引用格式: 何建强, 张玉萍, 滕志军. 基于 K-means 和改进 KNN 算法的风电功率短期预测系统[J]. 计算机测量与控制, 2022, 30(5): 156-162.

测。目前短期预测技术被广泛研究和应用, 主要研究方法有物理法、统计方法、机器学习法和空间相关性法。文献 [4] 对风电功率预测的物理方法和统计方法进行了综述。物理预测方法由于受风电场所处环境地形和地貌以及周围大气环境的影响比较大, 所以预测精度往往相对低。统计方法建模相对简单、计算速度快、预测成本低, 但是输入的预测参数单一, 不断地扩大预测时间尺度会降低预测精度。文献 [5-7] 对卡尔曼滤波法、时间序列法、最小二乘法、灰色预测法等常用的统计方法在风电功率预测中的应用进行研究, 但都存在易陷入局部最优的缺点。机器学习方法预测精度相对较高, 但预测模型比较复杂, 建模相对困难, 而且样本训练需要大量历史数据的支持, 训练速度慢。常用的方法有小波分析法, BP 神经网络法、SVM 等方法。文献 [8] 提出的人工神经网络法是风电功率短期预测当中最常用的预测方法, 具有非线性映射、自学习、容错性和泛化能力强等优点。但单一的预测方法具有不足之处, 为克服单一方法预测的不足, 组合预测法成为近年来风电功率预测研究的热点, 它是将一些统计方法和学习方法相结合, 提高预测精度。文献 [9-12] 提出了空间相关性法, 将其作为继物理模型、统计学之后的第 3 种模型, 它是一种通过邻近风电场气象信息对目标风电场进行预测的方法。文献 [13] 结合空间相关性的方法, 建立了预测模型, 解决了风速突变造成预测精度低的问题。但由于风速时变性较强, 预测的最大误差很难控制, 有时预测可能完全失效。

为解决单一模型在风电功率预测中的局限性, 本文针对文献 [14] 中提出的 KNN 算法在短期风电功率预测中的不足, 提出了基于 K-means 和改进 KNN 算法的风电功率预测, 该算法针对大量的历史样本集和预测样本集, 建模简单, 运算速度快, 预测准确率高。以吉林省某风电场 365 天的实际数据为样本, 对本文预测方法的优越性进行了分析验证。

## 1 KNN 算法概述

KNN 算法是一个经典、成熟的分类机器学习算法。它的基本原理为: 存在一个被称为特征空间的已知类别的样本集合, 如果某个样本与特征空间中的某  $K$  个样本相似或最邻近, 而且这  $K$  个最邻近的样本大多数是同一个类别, 那么此样本也被标记为该类别。重点强调的是特征空间中的每个样本必须已被正确分类。KNN 算法在本质上采用的是投票分类决策机制, 通过最邻近样本的类别决策待分类样本的类别, 决策前需要遍历整个样本空间, 决策时只依靠最近邻的少数样本。KNN 算法的优点是适合类域交叉或者大量重叠的待分类样本集。本文利用样本回归特性对样本进行功率预测, 即对于一个预测样本, 首先获取样本的  $K$  个最邻近已分类样本, 然后计算其属性的平均值, 将该平均值作为预测样本的属性值<sup>[15]</sup>。

## 2 改进的 KNN 算法

在大气边界层中, 风电功率历史样本数量庞大, 每个

预测样本针对不同的  $K$  值进行预测时, 每个样本对应每个不同的  $K$  值都要遍历一次特征空间, 那么随着特征空间的扩大, 就会成倍地增加运算量, 也会成倍的降低运算速度。同时在分类过程中, 如果特征空间样本类别不平衡时, 一部分样本容量较大, 另一部分样本容量较小, 随着  $K$  值的逐渐增大, 选取样本的误判率就会增大, 针对这一情况本文对 KNN 算法进行了改进<sup>[15]</sup>。

改进策略是: 同一个预测样本, 在同一个特征空间中, 针对不同的  $K$  值, 只需要遍历一次整体样本空间。首先提取预测样本的最大  $K$  值, 此时遍历整体样本空间, 提取  $K$  个最相似的样本, 重点是要对选取的  $K$  个样本按相似程度的从高到低进行排列, 并且将这  $K$  个相似样本保存到相似队列中。这样如果预测样本想提取  $K-1$  个相似样本的话, 就不需要再遍历整体特征空间了, 只需要从相似队列中提取前  $K-1$  个样本即可。这种改进策略是通过增加存储空间来加快预算速度。因为对于现在的计算机来说, 存储空间越来越大。假设一个风电场有 100 台风机, 每分钟记录一项风机的运行状态, 运行状态中包括 20 个属性, 那么 10 年的记录也只有 38.5 G。所以存储空间不是风电功率预测的压力。因此, 利用存储空间换取高效的预测空间是值得的。

改进的 KNN 算法包括两个过程, 第一个过程是计算预测样本选取最多近邻样本时对应的  $K$  值, 同时对获取的相似样本集合从高到低按照相似程度排序, 并存至优先级队列中; 第二个过程是当  $K$  值减小时, 从已存储的优先级队列中直接获取相似样本集, 根据相似距离决定相似权重, 从而用训练样本的加权功率值来计算预测样本的功率值。获取最长优先级队列的核心步骤如下。

第一步: 确定最近邻样本个数最多时对应的  $K$  值, 给出已知类别的训练样本集 (或者叫做特征空间)  $D$ ;

第二步: 设每一个预测样本为  $z = (x', y')$ , 开始执行 for 循环 for  $z = (x', y')$ ;

第三步: 计算  $z$  与每个训练样本  $(x, y) \in D$  间的相似距离  $d(x', x')$ ;

第四步: 选择与预测样本  $z$  最相似的  $K$  个训练样本  $D_z \subseteq D$ , 同时记录这  $K$  个样本与  $z$  的相似距离;

第五步: 按照与  $z$  相似程度对  $D_z$  从高到低排序, 将结果存至优先级队列 PriorityQueue $_z$  中;

第六步: end for; //结束 for 循环, 并保存好本次预测的优先级队列。

## 3 建立基于 K-means 和改进 KNN 算法的功率预测模型

### 3.1 预测原理

本文通过相似历史数据建立模型, 实现风电功率的预测。将风向、风速和风机变桨角度这 3 个参数作为输入参数, 选取不受时间序列影响而且样本集合的相似度比较高的数据作为样本数据。在输入参数后会自动寻找与之相似的样本数据集合, 并匹配对应的风电功率数据, 以此来对未来的输出功率进行推算, 预测模型框架如图 1 所示。最

后,通过吉林省某风电场实际数据来对预测模型的有效性和预测准确度进行检验。

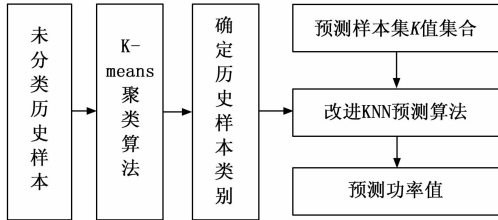


图 1 预测模型框架图

### 3.2 基于 K-means 的历史样本分类

由于本文所提改进的 KNN 算法通过相似历史数据进行预测时采用的数据分类方法与 K-means 聚类算法思想一致,因此在对历史样本进行分类时采用 K-means 算法,该方法在增加选中相似样本概率的同时,也保障了预测的准确性。该算法进行历史样本分类的过程如下。

Input:  $N$ //为区别改进 KNN 中的  $K$ , 此处用  $N$  表示划分子类的个数;  $\{Y_j\}$  //历史样本集, 其中包含  $j$  个对象。

Output:  $N$  个子类。

具体执行流程如下。

第一步:从历史样本集  $\{Y_j\}$  中随机选择  $N$  个样本作为初始分类的中心。

第二步:比较新读取样本对象和各个类中对象的均值,然后将其分配至最相似的类中。

第三步:对每个类的中心进行更新,重新对每个类中对象的均值进行计算。

第四步:until 每个类的中心不再变化结束。

其中:  $\{Y_j\}$  表示历史样本集合, 其中每个样本均包含风向、风速和风机变桨角度 3 个维度;  $N$  为将要划分子类的个数。

以吉林省某风电场 2019 年 2 月的历史数据为例,每隔一分钟记录一次风电机组的运行状态,一共 29 天,每台机组的历史样本数量为 41 760 条。以其中某一台的风机的历史样本记录为例,利用 Clementine 软件中的 K-means 聚类模型得出的聚类结果如 2 图所示。

### 3.3 基于改进 KNN 算法的功率预测

改进的 KNN 算法,是优化了相似历史样本集的搜索方式,从而提升了运算效率。按照时间复杂度进行计算,运行效率由  $n(k^2)$  提高为  $n(k)$ <sup>[16]</sup>。改进 KNN 的算法流程如图 3 所示。

算法的具体执行过程如下:

- 1) 初始化预测样本  $X_i$  和  $K$  值的集合。
- 2) 获取最大  $K$  值,赋予  $K_{max}$ ,  $K_i$  仅搜索  $K_{max}$  对应的相似历史样本集合。
- 3) 选取相似度计算公式。

本文在计算预测样本和历史样本的相似度时,选用适用于数值型数据的欧氏距离和曼哈顿距离计算公式,分别如公式 (1) 和公式 (2) 所示。

表	注解	风速(m/s)	风向	变桨...	功率(KW)	SKM-K-Means
1		5.176	4.492	6.330	192.000	聚类-1
2		5.060	10.405	7.493	204.000	聚类-1
3		4.712	-0.903	6.715	205.000	聚类-1
4		4.437	1.694	4.553	192.000	聚类-1
5		4.104	-1.147	4.304	197.000	聚类-1
6		5.916	18.202	9.735	86.000	聚类-1
7		5.959	9.604	8.381	86.000	聚类-1
8		4.567	20.958	8.503	87.000	聚类-1
9		3.756	-6.456	9.053	79.000	聚类-1
10		2.841	-1.911	8.346	85.000	聚类-1
11		4.785	-13.258	9.557	86.000	聚类-1
12		4.843	-7.796	7.420	82.000	聚类-1
13		3.944	5.408	10.2...	85.000	聚类-1
14		7.583	4.546	11.5...	85.000	聚类-1
15		4.901	13.987	10.1...	93.000	聚类-1
16		5.930	-3.550	10.5...	86.000	聚类-1
17		5.365	-3.282	12.9...	92.000	聚类-1
18		8.177	1.020	12.3...	169.000	聚类-1
19		8.206	-14.478	13.2...	91.000	聚类-1
20		8.134	5.825	12.2...	89.000	聚类-1
21		7.264	-2.087	12.2...	112.000	聚类-1
22		7.481	-2.338	11.9...	90.000	聚类-1
23		7.423	-8.434	88.9...	0.000	聚类-2
24		7.699	2.823	88.9...	0.000	聚类-2
25		7.003	7.361	88.9...	0.000	聚类-2
26		6.032	13.667	88.9...	0.000	聚类-2
27		6.046	-2.737	88.9...	0.000	聚类-2
28		7.177	-11.109	88.9...	0.000	聚类-2
29		8.423	-10.652	74.7...	0.000	聚类-2
30		6.220	0.563	88.9...	0.000	聚类-2
31		6.727	0.867	10.0...	198.000	聚类-1
32		4.451	21.487	4.491	389.000	聚类-1
33		6.655	-5.854	9.683	162.000	聚类-1
34		4.466	-12.111	0.000	564.000	聚类-1
35		6.191	19.425	3.618	191.000	聚类-1
36		4.625	8.568	1.906	190.000	聚类-1
37		4.466	16.597	5.848	204.000	聚类-1

图 2 某台风机的历史数据聚类结果

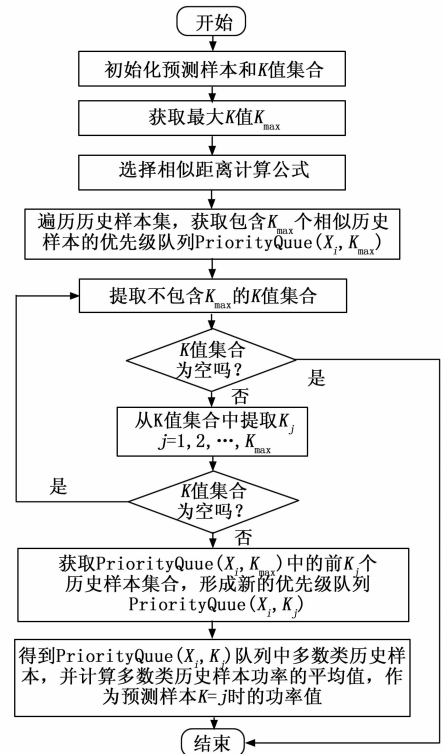


图 3 改进 KNN 算法流程

$$Dist_{Euc}(\vec{X}_i, \vec{Y}_j) = \left[ \sum_{z=1}^3 (x_{iz} - y_{jz})^2 \right]^{1/2} \quad (1)$$

$$Dist_{Mah}(\vec{X}_i, \vec{Y}_j) = \sum_{z=1}^3 |x_{iz} - y_{jz}| \quad (2)$$

式中,  $X_i$  和  $Y_j$  分别表示第  $i$  个预测样本和第  $j$  个历史样本。风向、风速和风机变桨角度为每个样本 3 个维度,  $z$  为维数。

4) 获取最大优先级队列  $PQ(X_i, K_{max})$ 。

在历史样本中随机选取  $K_{max}$  个样本, 按相似距离大排序, 生成优先级队列  $PQ(X_i, K_{max})$ 。遍历历史样本集, 并计算其与预测样本的相似距离。若相似距离比  $PQ(X_i, K_{max})$  中最后一个样本的相似距离大, 则继续遍历下一个历史样本, 否则移除  $PQ(X_i, K_{max})$  中最后元素, 将当前遍历样本值放入  $PQ(X_i, K_{max})$  尾部, 并按照相似程度重新自动排序, 遍历完所有历史样本, 得到最终  $PQ(X_i, K_{max})$ 。

5) 其余  $K$  值, 从最终  $PQ(X_i, K_{max})$  中提取对应的相似样本优先级子队列。如  $K = K_m$  时, 提取  $PQ(X_i, K_{max})$  中前  $m$  个历史样本, 作为  $K_m$  的  $PQ(X_i, K_m)$ 。

6) 计算  $PQ(X_i, K_m)$  中多数类别所有样本输出功率的平均值, 将其作为预测样本  $X_i$  的预测值。计算公式如公式 (3) 所示:

$$powerpr(X_i, K) = AVG(Y_{power}^{jC})$$

$$Y_{power}^{jC} \in \{Y_{power}^{1C}, Y_{power}^{2C}, \dots, Y_{power}^{jC}, \dots, Y_{power}^{KC}\} \quad (3)$$

式中,  $powerpr(X_i, K)$  表示预测样本  $X_i$  选取  $K$  个相似样本的功率值; 表示样本类别是  $C_z$  的第  $j$  个相似历史样本,  $z$  表示类型编号。AVG() 为计算多数类样本功率的加权平均值。

7) 计算所有  $K$  值下  $X_i$  对应的预测值, 结束预测。

最终功率预测值对应  $K$  值选取时, 要根据不同风机运转情况通过实验数据来进行确定最佳  $K$  值。

## 4 预测系统的设计

### 4.1 系统架构

本文采用 C/S 架构设计风电功率预测系统, 将预测模型转换为计算机可以完成的功能, 保障风电功率预测的高效性、稳定性和安全性。风电功率预测系统整体架构分为 3 层, 依次为基础数据层、功能服务层和用户层。每一层之间是相互独立且具有一定的联系, 风电功率预测系统架构图如图 4 所示。

#### 4.1.1 基础数据层

基础数据层又包括 3 部分。第一部分是风电实时数据的采集与整理; 第二部分是数据存储, 将整理完毕的实时数据, 存储于数据库中, 可以永久性的使用, 并定期维护; 第三部分建立数据访问接口。

#### 4.1.2 功能服务层

功能服务层主要是指风电功率预测系统需要完成的功能应用。本系统的应用服务主要有基础数据更新模块、实时数据展现模块、风电数据清洗模块、风电功率预测模块、预测误差分析模块、用户管理模块、系统设置模块、辅助模块等。

#### 4.1.3 用户层

用户层是与用户直接相关的部分。用户层是用户与系统进行交互的终端。用户可以输入相关

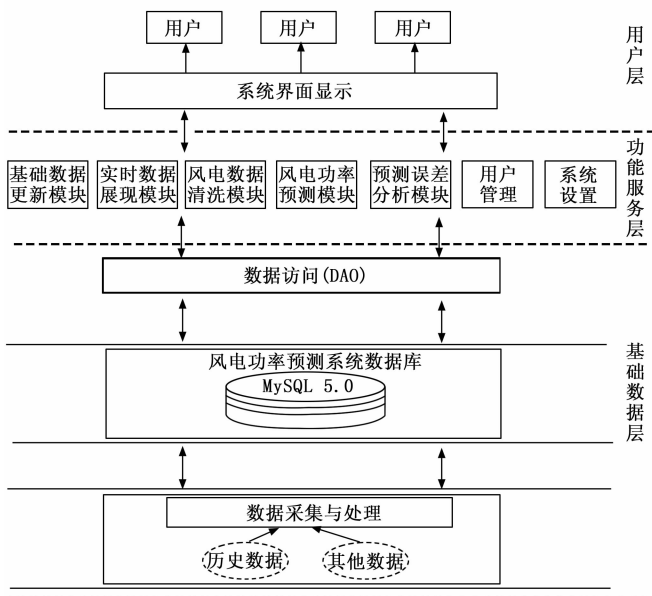


图 4 风电功率预测系统架构图

数据或者点击相关操作, 实现需要的功能。

### 4.2 系统功能

采用 java 语言实现风电功率预测系统的整体功能, 选用关系型数据库 MySQL 作为数据库管理系统。系统主要包括用户管理模块、核心模块和系统设置模块。系统整体功能模块图如图 5 所示。

用户管理模块主要包括用户注册模块、用户登录模块、信息修改模块等。用户注册模块主要是新用户在使用系统前进行用户注册。为了确定用户账号的安全性, 系统引入 MD5 加密技术, 对用户的密码以及密码修改提示答案采用密文的形式存储。这样可以防止非法用户的入侵, 保障系统的安全性。用户登录模块是任何一个软件系统的基础模块。该系统的用户登录模块需要用户通过多方面的认证合格, 才可以进入系统。主要包括权限认证、岗位认证、用户名、密码、验证码等合法验证。

用户修改模块主要方便用户对自身基本信息的更新。

系统设置模块包括系统用户权限管理、系统数据备份、系统数据还原。用户权限管理模块主要是根据风电功率预

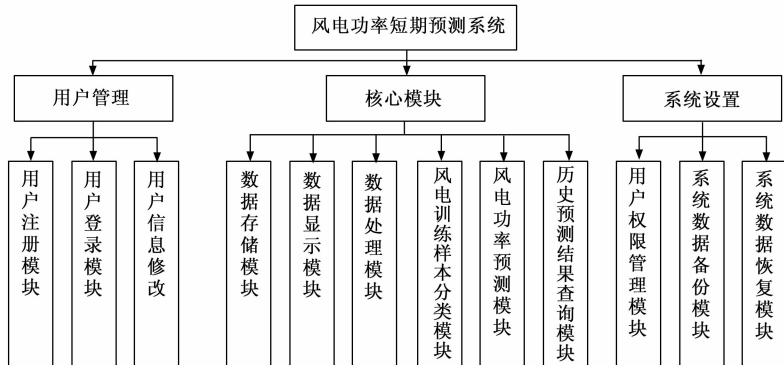


图 5 系统整体功能模块

测系统中用户的岗位类别、职称类别等判定用户的角色，并按照一定的安全策略为每种角色设置相应的权限。每个用户只可以访问自身权限范围内的内容。本系统的权限设置管理主要包括功能级权限管理和数据级权限管理。数据备份模块是系统容灾方案的一种。本系统考虑到电力系统属于专用网、风电数据的存储量大、不易恢复等特性，采用数据库备份方案。数据库备份即在备份机上创建数据库的拷贝版本，定期更新备份。系统数据恢复模块是指将丢失的或者异常数据恢复到正常的状态。本系统主要应用版本恢复将系统数据库中的数据恢复到前一个备份版本，使数据恢复正常。

### 4.3 核心模块设计

核心模块的主要功能为风电实时数据存储、实时数据查询、实时数据处理、风电训练样本分类、风电功率预测、误差分析和历史预测结果查询。该模块的 UML 活动图如图 6 所示。

风电实时数据存储模块主要用于获取风电场的实时数据。通过 I/O 读取程序，获取已经存储于记事本的风电功率实时采集数据。按照每一行为一个记录的方式，将获取的数据存入已经创建好的 MySQL 数据库，作为风电功率预测的数据源。

风电实时数据显示模块主要协助用户根据自己的意愿选择要浏览的风电数据。用户可以按照时间段模式进行预览。风电实时数据处理模块主要是清洗那些由于采集、传输和记录过程中导致的异常数据、缺省数据或者记录不完整的数据。利用相邻时间点数据的加权平均值修补缺省值和更正不完善的数据。

风电训练样本分类模块是由于预测模型需要输入具有确定类别的风电训练样本。该模块主要是将 K-means 聚类算法用 java 语言实现。

风电功率预测模块是系统的核心模块。该模块同样也是利用 java 语言实现预测算法的过程。该模块留有输入、输出两个接口。输入接口主要是输入训练样本和预测样本。输出接口就是输出风电功率的预测结果。

历史预测结果查询模块可以查询预测完毕的结果。一般用户可以通过预测时间段查询预测结果，也可以根据预测误差率或者准确率来查询预测结果。

## 5 实例验证

以吉林省某风电场 20 台机组的实时运行状态数据为数据基础，选取 2019 年 7 月 15 日到 2020 年 7 月 14 日的数据对算法进行验证。采样频率为每 5 分钟一次，抽取 300 天的

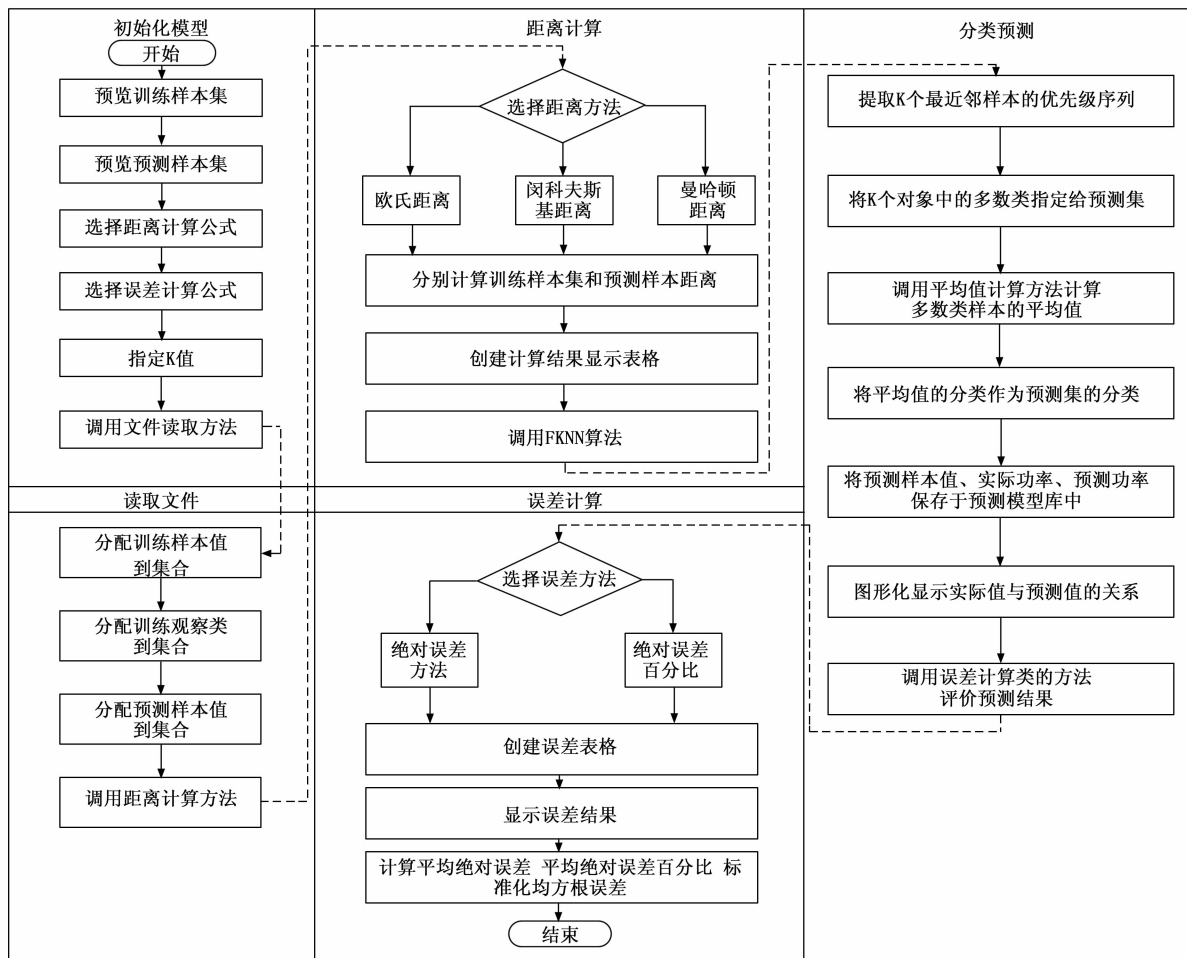


图 6 风电功率预测系统 UML 活动图

数据作为训练样本, 剩余 65 天的数据作为预测样本, 由于该风电每间隔 1 秒记录一次风机的实时运行状态, 但是实际预测中以秒为单位进行分割, 会产生大量的冗余数据, 因此本文按照每 5 分钟的采样频率记录历史数据, 将数据保存到数据库中。预测时间分辨率设定为 15 分钟, 将 24 小时划分为 96 个时间点进行预测。K 值的集合为 [1, 100], 分别利用曼哈顿距离和欧氏距离寻找相似样本。不同 K 值下两种距离计算方法的预测结果对比如 7 和图 8 所示。

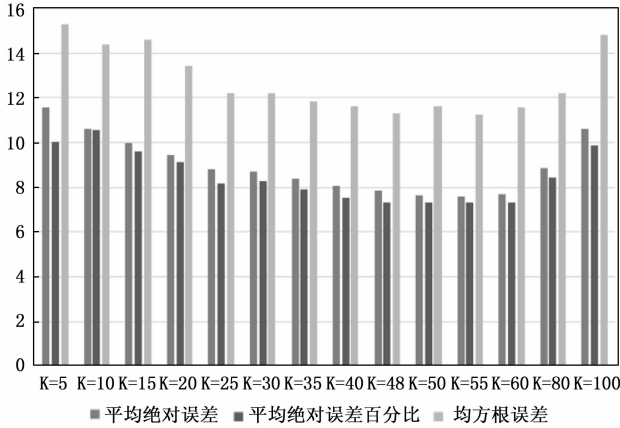


图 7 不同 K 值下曼哈顿距离计算方法的预测结果

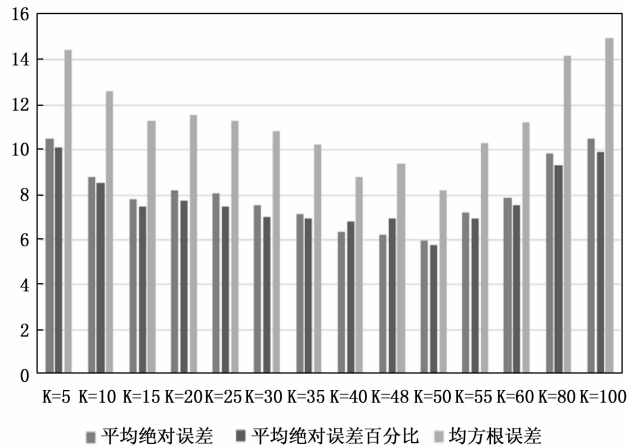


图 8 不同 K 值下欧氏距离计算方法的预测结果

由图 7 和图 8 可以看出, 当  $K=48$  时, 曼哈顿距离计算方法预测结果最好, 平均绝对误差最低。当  $K=50$  时, 欧氏距离计算方法预测结果较好, 预测误差也远低于国家规定范围。预测功率与实际功率值的对比结果分别如图 9 和图 10 所示。从图 7 和图 8 的预测误差率中可以看出, 预测误差随着 K 值的增加逐渐降低, 增加到一定值后, 随着 K 值的增加又逐渐升高。其与 K 值的变化关系如图 11 所示。通过大量的实验发现, 当 K 值在 45~55 时, 预测准确率最高。

本文将预测结果和文献 [17-18] 中的预测结果按照误差分析方法进行了对比, 预测误差对比结果如表 1 所示。

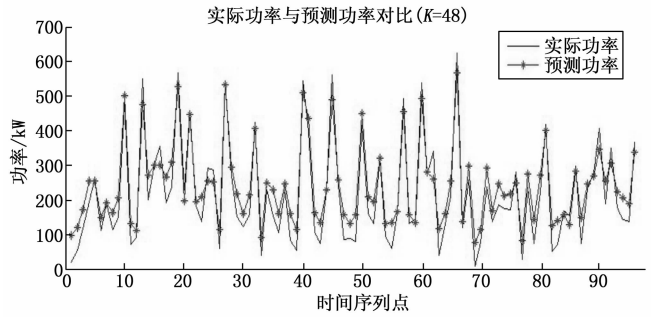


图 9 K=48 时未来 24 小时功率预测值与实际值对比

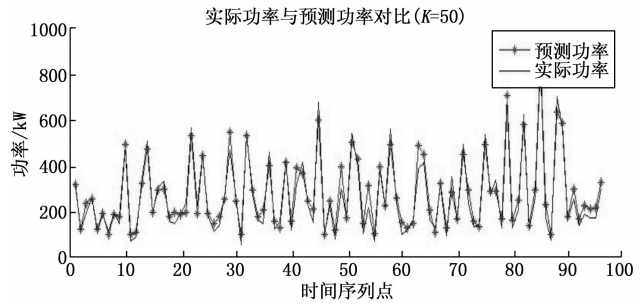


图 10 K=50 时未来 24 小时功率预测值与实际值对比

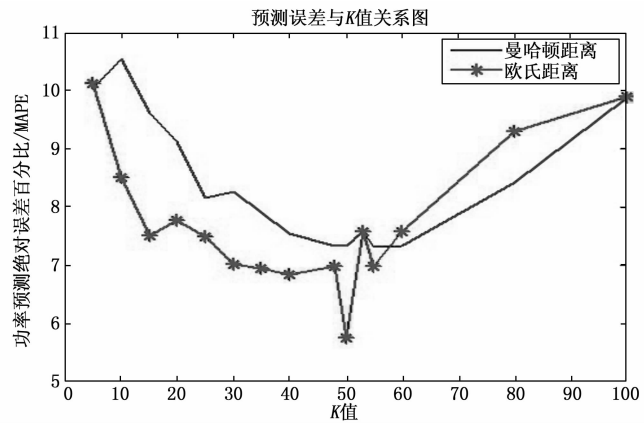


图 11 预测误差与 K 值的关系

从对比结果中可以看出采用本文的方法, 平均绝对误差和均方根误差预测误差率分别最大下降 1.08 和 0.48 个百分点。在相同的条件下, 本文的方法在预测时间最大提高了 5.45 个百分点。可见, 本文的预测方法提高了预测准确度。

表 1 预测误差对比

预测方法	平均绝对误差百分比/%	均方根误差/%
12.62	ANN	8.16
EMD-ANN	7.78	11.06
PCA-BP	12.8	18.9
SVM	10.6	18.7
RBF	12.4	15.6
FNN	17.4	24.4
改进 KNN 法(欧氏距离)	5.95	8.69
改进 KNN 法(曼哈顿距离)	7.83	7.36

改进的 KNN 预测模型在预测过程中不断地剔除训练样本中异常的或者影响预测准确度的样本,从而修正预测模型,使得预测模型的预测准确率不断自动地提升。因此,实验数据表明在对 20 台风机功率的预测中,误差率越来越低,每台风机的平均误差都小于 15%,满足目前风电功率预测系统的要求。

## 6 结束语

本文将 K-means 和改进的 KNN 算法应用到风电功率短期预测中,两种算法都是基于相似数据原理,学习方法不需要建立复杂的模型,预测模型简单、思想直观,更适合输入大量的历史样本数据,提高预测的准确率。采用 C/S 架构完成了风电功率预测系统的设计与实现,并以吉林省某风电场的历史数据为例,验证了该模型具有自学习、自修正能力,随着预测次数的增加,系统能够自动提升风电功率的预测准确度。该预测算法虽然具有一定的实用性,但如何取得最优 K 值,仍是下一步需要重点研究的内容。

### 参考文献:

[1] 王伟胜,王 铮,董 存,等. 中国短期风电功率预测技术现状与误差分析 [J]. 电力系统自动化, 2021, 45 (1): 17-27.

[2] 彭小圣,熊 磊,文劲宇. 风电集群短期及超短期功率预测精度改进方法综述 [J]. 中国电机工程学报, 2016, 36 (23): 6315-6326, 6596.

[2] 王彩霞,鲁宗相,乔 颖,等. 基于非参数回归模型的短期风电功率预测 [J]. 电力系统自动化, 2010, 34 (16): 78-82, 91.

[4] 谷兴凯,范高峰,王晓蓉,等. 风电功率预测技术综述 [J]. 电网技术, 2007 (S2): 335-338.

[5] 黄 慧,贾 嵘,董开松. 基于时空相关性的 NAR 动态神经网络风功率超短期组合预测 [J]. 太阳能学报, 2020, 41 (10): 311-316.

[6] 赵 欣,王爽心. 基于灰色关联和协整理论的风功率组合预测研究 [J]. 太阳能学报, 2017, 38 (5): 1299-1306.

[7] 修春波,任 晓,李艳晴. 基于卡尔曼滤波的风速序列短期预测方法 [J]. 电工技术学报, 2014, 29 (2): 253-259.

[8] 范高峰,王伟胜,刘 纯. 基于神经网络的风电功率短期预测系统 [J]. 电网技术, 2008 (22): 72-76.

[9] 薛禹胜,陈 宁,王树民,等. 关于利用空间相关性预测风速的评述 [J]. 电力系统自动化, 2017, 41, (10): 161-169.

[10] 杨正瓩,冯 勇,熊定方,等. 基于季风特性改进风电功率预测的研究展望 [J]. 智能电网, 2015, 3 (1): 1-7.

[11] AHMED A, KHALID M. A review on the selected applications of forecasting models in renewable power systems [J]. Renewable and Sustainable Energy Reviews, 2019, 100: 9-21.

[12] YANG B, ZHONG L, WANG J, et al. State of the art one-stop handbook on wind forecasting technologies: An overview of classifications, methodologies, and analysis [J]. Journal of Cleaner Production, 2021, 283: 124628.

[13] 胡 帅,向 月,沈晓东,等. 计及气象因素和风速空间相关性的风电功率预测模型 [J]. 电力系统自动化, 2021, 45 (7): 28-36.

[14] 吴问足,乔 颖,鲁宗相,等. 风电功率概率预测方法及展望 [J]. 电力系统自动化, 2017, 41 (18): 167-175.

[15] 朱念芳,林善明. 基于邻域 KNN 算法的风电功率短期预测模型 [J]. 电测与仪表, 2017, 54 (16): 20-24.

[16] 郭晓利,张玉萍,曲朝阳,等. 基于 FKNN 算法的风电功率短期预测 [J]. 电测与仪表, 2014, 51 (15): 1-7.

[17] 杨 茂,董骏城,齐 玥. 基于近似熵的风电功率可预测性研究 [J]. 太阳能学报, 2016, 37 (10): 2710-2718.

[18] 周松林,茆美琴,苏建徽. 基于主成分分析与神经网络的风电功率预测 [J]. 电网技术, 2011, 35 (9): 128-132.

[15] 王勇军,李 智,孙山林,等. 轻小型无人机遥感组网飞行的高程安全监测冗余容错算法研究 [J]. 地球信息科学学报, 2019 (4): 542-551.

[16] 刘丙涛,周 炎,祖家奎. 无人直升机跟踪地面移动目标实时仿真系统设计与实现 [J]. 计算机应用与软件, 2019, 36 (1): 22-27.

[17] 谢 檬,宋 菲. 基于 LabVIEW 的无人机航线控制系统的设计 [J]. 自动化与仪表, 2019, 34 (8): 43-46.

[18] 王文亮,马 磊,韩 蕾. 无人机综合仿真与测试平台技术研究 [J]. 计算机测量与控制, 2019, 27 (9): 13-17.

[19] 赵广磊,高儒帅,陈健楠. 具有执行器故障的四旋翼无人机自适应预定性能控制 [J]. 控制与决策, 2021, 36 (9): 2103-2112.

[20] 卢海松,牛洪海,蔡 丹,等. 大型调相机除盐水无人值守控制系统设计与应用 [J]. 自动化仪表, 2020, 41 (4): 89-93.

[21] 钟 赞,姚佩阳,张杰勇,等. 有人/无人机协同作战系统 C2 结构和行动计划适应性设计方法 [J]. 空军工程大学学报 (自然科学版), 2019, 20 (3): 38-45, 102.

(上接第 142 页)

[8] 通用原子公司公布“雀鹰”小型无人机 [EB/OL]. (2020-10-14). [2021-11-23]. <http://3g.163.com/dy/article/FOS-8V4TE0528QLHI.html>.

[9] 美军成功用无人机投放小型无人机 [EB/OL]. (2021-04-06). [2021-11-23]. <http://www.guancha.cn/military-affairs/2021-04-06-586443.shtml>.

[10] 周 强,夏明赞. 基于改进 SSD 算法的多无人机检测研究 [J]. 信息技术, 2020, 44 (12): 71-76.

[11] 沈明松. 基于视觉识别的输电杆塔无人机自动巡检系统设计 [J]. 信息技术, 2020, 44 (9): 148-152.

[12] 张 亮,方 圆,蔡梦臣,等. 基于能量消耗 MPR 的无人机 OLSR 路由协议评估 [J]. 信息技术, 44 (9): 157-160.

[13] 利 锋,杨中柱,黄祖胜,等. 结合混合粒子群算法的植保无人机航线设计方法 [J]. 小型微型计算机系统, 2020, 41 (9): 1826-1832.

[14] 潘 影,张燕杰,武俊喜,等. 基于遥感和无人机数据的草地 NDVI 影响因子多尺度分析 [J]. 草地学报, 2019, 27 (6): 1766-1773.