

文章编号:0254-0096(2019)12-3578-08

# 基于熵权综合关联度数据挖掘的 风电功率实时预测研究

杨 茂<sup>1</sup>, 马 剑<sup>1</sup>, 李大勇<sup>2</sup>, 苏 欣<sup>3</sup>, 孙 涌<sup>4</sup>

(1. 现代电力系统仿真控制与绿色电能新技术吉林省重点实验室(东北电力大学), 吉林 132012;  
2. 国网吉林省电力有限公司通化供电公司, 通化 130022; 3. 东北电力大学理学院, 吉林 132012;  
4. 国网淄博供电公司, 淄博 255000)

**摘 要:** 采用一种基于熵权综合关联度指标, 来量化评估不同历史时段风电样本与待预测时段参考样本间的复杂非线性映射关系, 解决预测模型输入与输出变量的相关性冗余问题, 并与 Pearson 相关系数、Kendall 相关系数、Spearman 相关系数、互信息相关系数指标进行对比研究。然后, 采用一种神经网络改进模型, 通过相似样本筛选、隐含层结构寻优、网络权重赋初值等, 减小训练数据冗余度的影响及加快收敛速度, 提高预测模型的泛化能力和计算效率。通过吉林省某风电场的实测数据进行实验仿真, 结果表明该方法可有效提高风电功率实时预测准确率。

**关键词:** 关联信息挖掘; 熵权综合关联系数; 相关性冗余; 模型泛化能力

**中图分类号:** TM614

**文献标识码:** A

## 0 引 言

社会发展对能源的需求以及现在所面临的环境问题, 使得各个国家对风能、太阳能等清洁能源的开发利用逐渐重视<sup>[1]</sup>。中国目前的风力发电事业正在高速向前发展。但风电大规模接入电网会给电力系统的安全运行带来一些新问题, 其中最显著的问题是对电力系统运行调度的影响<sup>[2]</sup>。因风的波动性与间歇性对电能质量存在威胁, 给风电并网带来不利影响, 使得风速和风电功率的准确预测成为解决这些问题的重要手段<sup>[3-6]</sup>。实际工程应用中的风电功率预测方式有 2 种, 即日预测和实时预测。日预测是预测未来一天, 即次日 00:00~24:00 的风电功率; 实时预测是从预测上报时刻起未来 15 min 至 4 h 的风电功率。

一般利用历史风电数据进行预测建模时需要先对历史风电数据进行去噪<sup>[7]</sup>和降维<sup>[8]</sup>处理, 然后将处理后的数据作为建模数据进行风电功率预测。风电历史数据样本空间巨大, 且存在一定关联性, 若直接将原始数据作为模型的输入会影响预测

的精度和稳定性。基于风电数据变化趋势具有一定相似的特点, 一些学者提出基于相似理论的聚类分析(cluster analysis, CA)方法来预测风速<sup>[9]</sup>。持续性法、时间序列法、神经网络法、卡尔曼滤波等<sup>[10-13]</sup>均是风电功率预测常用的方法。其中的神经网络法在非线性和泛化能力方面具有显著的优势, 能够较好地描述风电功率复杂非线性。但输入变量和训练样本的选取会直接影响神经网络模型的预测效果<sup>[14]</sup>。

对风电功率历史数据的关联信息进行挖掘, 能有效提高风电功率实时预测精度及计算效率。大多数模型的输入变量选取采用根据相关性人为设定阈值的分析方法<sup>[15]</sup>, 这种选取方法不可避免带来个人主观因素, 造成信息冗余或损失, 且影响风电功率的因素是多方面的, 若将全部因素作为模型输入会导致神经网络的训练负担加重, 预测精度降低<sup>[14]</sup>。已有实验证明, 选取相似性较高的数据训练模型可有效提高训练效率以及模型的泛化能力<sup>[16]</sup>。

基于以上研究, 本文提出一种熵权综合关联度指标, 对大量的风电功率历史数据进行挖掘, 找出

收稿日期: 2017-05-16

基金项目: 国家自然科学基金(51307017); 吉林省产业技术与专项开发项目(2014Y124); 国家重点研发计划(2018YFB0904203)

通信作者: 杨 茂(1982—), 男, 博士、教授, 主要从事风力发电技术方面的研究。yangmao820@163.com

关联性较强的相似样本数据,并与 Pearson 相关系数、Kendall 相关系数、Spearman 相关系数、互信息相关系数指标进行对比研究。然后通过对输入样本进行筛选、网络权重赋值及隐含层结构寻优对人工神经网络模型进行优化。通过对吉林省某风电场的实测数据进行仿真实验,验证本文所提方法在风电功率实时预测准确率提高方面的有效性。

## 1 熵权综合关联度指标

### 1.1 常见的关联度评估指标

常用评价变量间相关性的指标主要有 4 种: Pearson 相关系数、Kendall 相关系数、Spearman 相关系数、互信息相关系数。

Pearson 相关系数即线性相关性系数,其计算公式如式(1)所示:

$$\rho_{x,y} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \sigma_y} \quad (1)$$

式中,  $\text{cov}(X,Y)$  ——  $X$  和  $Y$  的协方差;  $\sigma_x$ 、 $\sigma_y$  ——  $X$  和  $Y$  的标准差;  $\mu_x$ 、 $\mu_y$  ——  $X$  和  $Y$  的期望  $E(X)$  和  $E(Y)$ 。

Kendall 相关系数,即和谐系数,其计算方法为:对于随机变量  $X$  和  $Y$  的 2 对观察值  $X_i$ 、 $Y_i$  和  $X_j$ 、 $Y_j$  ( $X_i$  和  $X_j$  分别表示随机变量  $X$  的第  $i$  次和第  $j$  次观察值,  $Y_i$  和  $Y_j$  分别代表随机变量  $Y$  的第  $i$  次和第  $j$  次观察值),若  $X_i \leq Y_i$  且  $X_j \leq Y_j$ , 或  $X_i > Y_i$  且  $X_j > Y_j$ , 则称这 2 对观察值为和谐,否则为不和谐。

Kendall 相关系数  $\tau$  的计算公式:

$$\tau = \frac{n_1 - n_2}{\frac{1}{2}n(n-1)} \quad (2)$$

式中,  $n_1$  —— 和谐对数;  $n_2$  —— 不和谐对数。总对数为  $\frac{1}{2}n(n-1)$ 。

Spearman 相关系数,即秩相关系数。其值仅与 2 个相关变量值之间的大小关系有关,与其具体值无关。在 Pearson 系数的方法中,用原随机变量中的等级顺序替换原始数据即得到 Spearman 相关系数的计算方法。

互信息表征多个随机变量之间相互包含信息的多少。随机变量  $X$  随机波动的不确定度可用信

息熵  $H$  进行量化评估:

$$H(X) = -\sum_{i=1}^N p(X_i) \log_2(p(X_i)) \quad (3)$$

若将式(3)扩展至二维随机变量  $X$  和  $Y$  范畴,由香农信息论可得到联合熵  $H(X,Y)$ 、条件熵  $H\left(\frac{X}{Y}\right)$ 、互信息  $I(X,Y)$ ,如式(4)~式(6)所示:

$$H(X,Y) = -\sum_{i=1}^N \sum_{j=1}^M p(X_i Y_j) \log_2(p(X_i Y_j)) \quad (4)$$

$$H\left(\frac{X}{Y}\right) = -\sum_{i=1}^N \sum_{j=1}^M p\left(\frac{X_i}{Y_j}\right) \log_2\left(p\left(\frac{X_i}{Y_j}\right)\right) \quad (5)$$

$$I(X,Y) = \sum_{i=1}^N \sum_{j=1}^M p(X_i Y_j) \log_2\left(\frac{p(X_i Y_j)}{p(X_i)p(Y_j)}\right) \quad (6)$$

式中,  $p(X_i Y_j)$  —— 是随机变量  $X$  取  $X_i$  值且随机变量  $Y$  取  $Y_j$  值时的联合概率;  $N$  —— 随机变量  $X$  最大取值数量;  $M$  —— 随机变量  $Y$  最大取值数量。

互信息  $I(X,Y)$  表示随机变量  $X$  和  $Y$  之间共享信息量的多少,与信息熵、条件熵、联合熵之间的关系如式(7)所示<sup>[17]</sup>:

$$\begin{aligned} I(X,Y) &= H(X) - H\left(\frac{X}{Y}\right) \\ &= H(X) - [H(X,Y) - H(Y)] \\ &= H(X) + H(Y) - H(X,Y) \end{aligned} \quad (7)$$

### 1.2 熵权关联度

熵权法是通过指标变异性大小来确定各指标客观权重的。一般来说,指标值变异程度越大说明其包含信息量愈多,则其对应的信息熵愈小,指标权重愈大,表明在综合评价中所能起到的作用也越大<sup>[18]</sup>。同理,指标值变异程度越小说明其包含信息量愈少,则其对应的信息熵愈大,指标权重愈小,表明在综合评价中所能起到的作用也越小。设风电历史数据的总时段数为  $M$ ,  $G_i = \{g_{i1}, g_{i2}, g_{i3}, \dots, g_{ij}, \dots, g_{iM}\}$  为历史时段与待预测时段关联度指标的定义,其中  $G_i$  中元素数量为  $G_m^2$  个(本文的分指标数为 4,故  $l=1,2,3,4$ )。

1)对历史时段与待预测时段关联度进行标准化处理:

$$Y_i = \frac{G_i - \min(G_i)}{\max(G_i) - \min(G_i)} \quad (8)$$

式中,  $Y_i$ ——关联度标准化之后的值;  $G_i$ ——历史时段与待预测时段关联度。

2) 求各指标数据对应的信息熵  $E_k$  :

$$E_k = -\ln(C_M^2)^{-1} \sum_{m=1}^{C_M^2} p_m \ln(p_m) \quad (9)$$

式中,  $p_m = \frac{Y_m}{\sum_{m=1}^{C_M^2} Y_m}$ ,  $p_m$  为  $p$  的第  $m$  次取值, 其中  $Y_m$

为关联度标准化之后的取值。若  $p_m = 0$ , 则定义  $\lim_{p_m \rightarrow 0} p_m \ln p_m = 0$ 。

3) 确定  $K$  个指标对应的权重:

$$w_k = \frac{1 - E_k}{K - \sum E_k} \quad (10)$$

4) 根据各指标熵权的大小, 对其进行加权组合, 即可得熵权综合关联度指标  $d$  :

$$d = \sum_{k=1}^4 w_k d_k \quad (11)$$

式中,  $d_k$ ——各指标对应的值;  $w_k$ ——各指标对应的权重。

## 2 模型输入和输出变量之间的关联度评估

本文研究主要是将不同历史时段风电功率数据, 与待预测时段, 即未来 4 h, 16 个数据点的风电信息进行关联度分析, 筛选出强相关的历史数据作为预测模型输入变量(简称相似样本)。由于实际工程中待预测时段风电功率为未知量, 因此本文采用数值天气预报(numerical weather prediction, NWP)数据来对待预测时段参考样本进行构造。

### 2.1 待预测时段参考样本的确定

国内近年来的测风系统日趋完备, 数字天气预报, 即 NWP 预测精度得到一定提高<sup>[19]</sup>。NWP 数据一般包括风速、气温、气压、湿度等, 这些因素均会对风电功率产生一定影响。本文首先对历史风电功率和历史 NWP 数据进行相关性的挖掘, 实验结果如表 1 所示, 结果表明风速是风电功率预测中最重要的影响因子, 因此利用 NWP 风速数据作为神经网络模型的输入, 风电功率作为模型的输出, 进行风-电转换后的风电功率数据作为预测时段参考样本。

表 1 NWP 数据与风电功率相关性统计

相关系数	风速	风向	气压	温度	湿度
Pearson 系数	0.479	0.013	0.295	0.006	0.246
Spearma 系数	0.519	0.078	0.249	0.116	0.232

### 2.2 基于熵权综合关联度的风电关联信息挖掘

Pearson 相关系数(即线性相关系数)是从波形相似度来判断随机变量相关性; 秩相关系数是从变化趋势一致性来判断随机变量之间相关性; 互信息关联度是从随机变量概率密度分布特性来判断随机变量之间的相关性。由于风电功率具有波动性和随机性, 较难发现其规律性, 且采用单一的关联度指标不能准确刻画风电功率的相关性。因此本文运用熵权综合关联度方法, 对风电功率历史数据的关联性进行挖掘。通过熵权法对上述各指标进行加权组合, 可综合反映历史数据的局部波形相似及变化趋势和概率密度分布特性的整体关联程度。

根据上述分析, 利用待预测时段的 NWP 风速数据, 进行风-电转换后得到的风电功率数据作为熵权综合关联度指标的参考基准, 将历史时段风电功率数据与其两两进行比较, 得到熵权综合关联度指标, 按照指标大小进行优先级排序, 选取关联度较大的前  $j$  个样本作为预测模型的输入信息(本文中取  $j=5$ , 远小于原始数据时段数), 其流程图如图 1 所示。该方法能对历史相似样本进行快速识别, 有

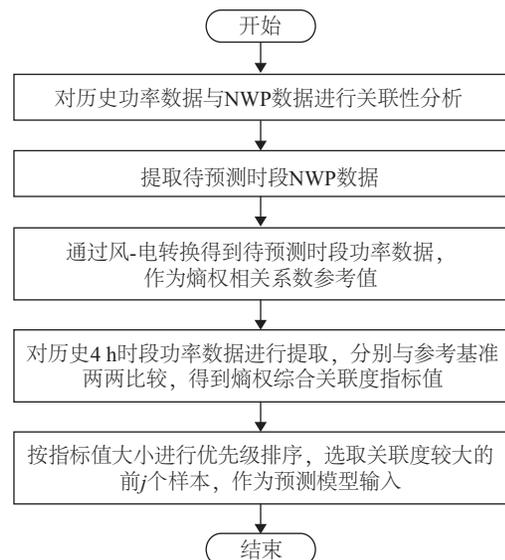


图 1 基于熵权综合关联度的风电关联信息挖掘流程图

Fig. 1 Flow chart of wind power correlation information mining based on entropy weight

效过滤掉相关性弱或无关数据的干扰,实现对风电历史数据的有用信息挖掘。

### 3 神经网络优化预测模型

将上述风电历史数据关联信息挖掘算法与风电功率实时预测模型相结合,对人工神经网络模型进行优化,其基本框架如图2所示,图中 $P$ 为历史风电功率序列, $P_1, P_2, P_3, \dots, P_n$ 为各时段风电功率序列。

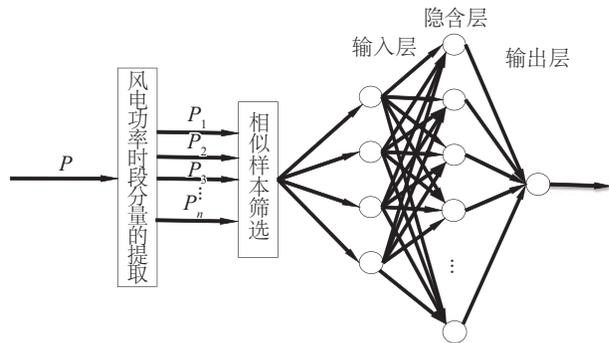


图2 神经网络优化模型的基本框架

Fig. 2 Basic frame of neural network optimization model

优化的具体步骤如下:

1)提取风电功率数据时段分量:将每15分钟1个点的历史风电功率数据按照每4小时16个点的时段集合 $\{P_1, P_2, \dots, P_j\}$ 进行提取,其中 $t$ 为历史风电功率数据时段数。2)筛选相似样本:将所提取历史风电功率时段集合样本归一化,然后分别计算其与待预测时段参考样本的熵权综合关联度,以其大小排序为标准,选取前 $j$ 个相关性较强的样本作为相似样本(即 $\{P_1, P_2, \dots, P_j\}$ ,本文 $j$ 取5),作为人工神经网络模型的输入变量。3)隐含层网络结构优化:采用3层人工神经网络结构,输入层为16个节点,输出层为1个节点,隐含层采用单层结构。本文采用网络结构增长寻优方式,节点数由10个递增至30个,分别计算测试数据的预测误差,选取误差最小的节点数。4)网络权重赋初值:连接权重与阈值是影响神经网络的训练速度和收敛速率的2个重要因素,其初值通常采用随机赋值方式,这样易出现输出结果不稳定的情况。本文将相似样本对应的关联度作为人工神经网络输入层的初始权值 $w_1$ ,初始阈值 $b_1$ 设为0,有利于模型收敛性和稳定性的提高。

### 4 算例分析

为验证本方法的有效性,文中对吉林省西部风电场A在2014年4~5月份和风电场B在2014年5~6月份实测数据进行分析<sup>[20]</sup>,风电场A总装机容量为400.5 MW,风电场B总装机容量为49.5 MW,单台机组额定输出功率为1.5 MW。文中主要对实时滚动预测1天和1个月的结果进行分析。文中风电功率实时预测的评价指标采用的是国家能源局对并网风电场的预测预报考核指标<sup>[21]</sup>。

1)日平均预测计划曲线准确率 $r_1$ :

$$r_{1i} = \left[ 1 - \sqrt{\frac{1}{16} \sum_{k=1}^{16} \left( \frac{P_{Mi}^k - P_{Pi}^k}{P_{cap}} \right)^2} \right] \times 100\% \quad (12)$$

$$r_1 = \frac{1}{96} \sum_{i=1}^{96} r_{1i} \quad (13)$$

式中, $r_{1i}$ ——第 $i$ 步实时预测的准确率。 $P_{Mi}^k$ ——第 $i$ 步实时预测中第 $k$ 时刻的实际风电功率; $P_{Pi}^k$ ——第 $i$ 步实时预测中第 $k$ 时刻的风电功率预测值; $P_{cap}$ ——风电场总装机容量。

2)日平均预测计划曲线合格率 $r_2$ :

$$r_{2i} = \frac{1}{16} \sum_{k=1}^{16} B_i^k \times 100\% \quad (14)$$

$$\left[ 1 - \left| \frac{P_{Mi}^k - P_{Pi}^k}{P_{cap}} \right| \right] \times 100\% \geq 85\%, B_i^k = 1 \quad (15)$$

$$\left[ 1 - \left| \frac{P_{Mi}^k - P_{Pi}^k}{P_{cap}} \right| \right] \times 100\% < 85\%, B_i^k = 0 \quad (16)$$

$$r_2 = \frac{1}{96} \sum_{i=1}^{96} r_{2i} \quad (17)$$

式中, $r_{2i}$ —— $i$ 步实时预测的合格率; $B_i^k$ ——第 $i$ 次预测结果合格指标。

3)全天预测结果均方根误差 $r_3$ :

$$r_3 = \sqrt{\frac{1}{96 \times 16} \sum_{i=1}^{96} \sum_{j=1}^{16} \left( \frac{P_{Mi}^k - P_{Pi}^k}{P_{cap}} \right)^2} \times 100\% \quad (18)$$

将本文所提熵权综合关联度指标与Pearson相关系数、Kendall相关系数、Spearman相关系数、互信息相关系数进行对比分析,分别将其用于人工神经网络优化模型输入样本的筛选,利用提出的人工神经网络优化模型进行风电功率实时预测研究,用日平均预测计划曲线准确率( $r_1$ )、日平均预测计划曲线合格率( $r_2$ )和全天预测结果均方根误差( $r_3$ )来评价模型的预测效果,并采用计算时间评价模型

计算效率。本文以4月1~19日的的数据作为预测模型的训练数据,20日的的数据作为待预测的数据进行验证。

首先,将4月1~19日的的风电功率数据按时段分量进行提取,得到114个时段分量;其次,利用4月1~19日的NWP数据对4月20日的的风电功率前16个点进行预测,得到待预测时段的参考样本;然后,分别求取各时段分量与待预测时段参考样本的Pearson相关系数、Kendall相关系数、Spearman相关系数、互信息关联系数,以及熵权综合关联系数;按照各相关系数大小进行排序,取相关性强的前 $j$ 个时段分量作为预测模型的输入(本文中 $j=5$ ),进行模型训练。利用不同关联度筛选的相似时段分量作为神经网络模型输入,以对4月20日的的风电功率进行实时预测为例,将00:00时刻~04:00时刻16个点的功率值作为输入信息,预测04:15时刻的功率值,然后将00:15~04:15时段16个点的功率值作为模型输入,预测04:30时刻功率值,以此类推进行滚动预测,其预测结果评价指标统计如表2所示。图3为采用熵权综合关联系数最好和最坏预测结果。

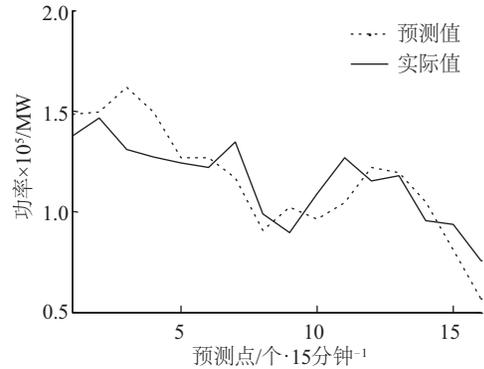
表2 风电场A4月20日不同关联度指标对应的预测效果

Table 2 Prediction results of wind farm A in different correlation index on April 20th

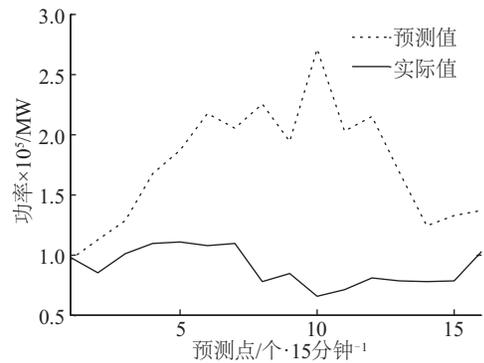
不同关联度指标	$r_1$	$r_2$	$r_3$	计算时间/s
未采用关联度挖掘	0.836	0.614	0.179	74.571
Pearson 相关系数	0.864	0.721	0.154	48.865
Kendall 相关系数	0.865	0.740	0.161	48.325
Spearman 相关系数	0.890	0.785	0.131	48.771
互信息相关系数	0.890	0.787	0.130	48.676
熵权综合关联系数	0.903	0.871	0.118	48.623

由表2可得,对预测模型的输入采用关联度数据挖掘与未采用的情况相比较,采用Pearson相关系数、Kendall相关系数、Spearman相关系数和互信息相关系数,以及熵权综合关联系数对输入样本进行筛选后,计算效率得到显著提高,计算消耗时间减少,准确率得到明显提高,采用本文所提方法准确率相比未采用关联度数据挖掘提高了6.73%,比采用互信息相关系数对输入样本进行挖掘的准确率提高了1.32%。此外,在合格率和均方根误差

也获得较优的效果,采用熵权综合关联系数对应的预测效果更好,合格率达到87.11%。



a. 最好预测结果



b. 最坏预测结果

图3 风电场A在4月20日最好与最坏预测结果图

Fig. 3 Best and worst prediction results of wind farm A on April 20th

同理,对风电场A5月和风电场B在5月20日及其6月份的风电功率进行实时预测,其预测结果评价指标统计结果如表3所示。由表3可得采用熵权综合关联系数对应的预测效果最好,准确率均达到87%以上,且计算时间显著减少。在表3中预测风电场B在5月20日的的风电功率准确率达到88.88%,均方根误差为0.1322,合格率为80.86%。仅采用Pearson相关系数对历史数据关联度进行挖掘时,预测模型的准确率相比未采用关联度挖掘低了0.41%,此时Pearson相关系数不能很好地挖掘历史数据的相关性。而采用Kendall相关系数、Spearman相关系数和互信息相关系数对历史数据关联度进行挖掘,预测效果较好,本文所提方法能有效避免某一关联度指标失效的情况,同时也证明了本文所提方法的有效性和可行性。

表3 风电场A在5月份、风电场B在5月20日、风电场B在6月份不同关联度指标对应的预测效果

Table 3 Prediction results of wind farm A in different correlation index in May, wind farm B in different correlation index on May 20 and in June

不同关联度 指标	5月份(风电场A)				5月20日(风电场B)				6月份(风电场B)			
	$r_1$	$r_2$	$r_3$	计算 时间/s	$r_1$	$r_2$	$r_3$	计算 时间/s	$r_1$	$r_2$	$r_3$	计算 时间/s
未采用关联度挖掘	0.830	0.186	0.685	2143.252	0.867	0.154	0.717	73.952	0.834	0.184	0.695	2123.562
Pearson 相关系数	0.855	0.163	0.664	1424.518	0.863	0.170	0.766	51.674	0.851	0.156	0.632	1422.952
Kendall 相关系数	0.858	0.171	0.741	1479.537	0.875	0.142	0.740	51.667	0.856	0.176	0.740	1553.633
Spearman 相关系数	0.843	0.185	0.666	1476.994	0.876	0.150	0.768	51.513	0.856	0.179	0.659	1507.135
互信息相关系数	0.860	0.160	0.750	1465.515	0.876	0.145	0.772	51.703	0.870	0.148	0.777	1457.392
熵权综合关联系数	0.879	0.143	0.804	1467.452	0.889	0.132	0.819	51.865	0.875	0.146	0.810	1467.452

## 5 结 论

预测模型的输入样本对模型的训练、计算效率及预测精度有重要影响,因此,对输入样本进行关联度挖掘显得非常必要。本文提出基于熵权综合关联度评价指标对风电功率历史数据进行挖掘,筛选出关联度较高的数据样本作为模型输入,并利用人工神经网络优化模型进行风电功率实时预测,可有效解决传统模型训练数据冗余度大和收敛速度慢等问题,提高预测模型的鲁棒性和泛化能力。通过风电场的实测数据进行仿真实验,分析结果表明,本文提出的风电功率实时预测方法,与未进行数据挖掘的预测方法相比,对未来一天预测结果的准确率提高了 6.73%,合格率达到 87.11%,均方根误差得到显著下降,对未来一个月预测结果准确率达到 87.45%,合格率达到 80.93%。此外,在 NWP 数据不完整的情况下,如何保证数据挖掘的有效性,还需进一步研究。

### [参考文献]

- [1] 刘 波,贺志佳,金 昊. 风力发电现状与发展趋势[J]. 东北电力大学学报, 2016, 36(2): 7—13.
- [2] Liu Bo, He Zhijia, Jin Hao. Wind power status and development trends[J]. Journal of Northeast Electric Power University, 2016, 36(2): 7—13.
- [2] 许玉贞,詹红霞,张瑞冬,等. 考虑风电预测误差成本的微网经济调度[J]. 电气技术, 2015, (5): 26—29, 34.
- [3] Xu Yuzhen, Hong Xia Zhan, Zhang Ruidong, et al. Consider wind power prediction error cost of micro

economic net scheduling[J]. Electric Technology, 2015, (5): 26—29, 34.

- [3] 杨德友,蔡国伟. 基于因散经验模式分解与最小二乘支持向量机的风电场短期风速预测[J]. 东北电力大学学报, 2015, 35(3): 44—49.
- [3] Yang Deyou, Cai Guowei. The short-term wind speed forecasting for wind farm based on EEMD and LS-SVM [J]. Journal of Northeast Electric Power University, 2015, 35(3): 44—49.
- [4] 杨 茂,季本明. 基于局域一阶加权法的风电功率超短期预测研究[J]. 东北电力大学学报, 2015, 35(5): 6—10.
- [4] Yang Mao, Ji Benming. The Ultra-short-term forecasting of wind power based on local one-order weighted method [J]. Journal of Northeast Electric Power University, 2015, 35(5): 6—10.
- [5] 刘红柳,杨 茂. 基于小波变换和支持向量机的风电功率爬坡事件识别与预测[J]. 东北电力大学学报, 2016, 36(6): 30—35.
- [5] Liu Hongliu, Yang Mao. An Approach for wind power climbing event identification and prediction based on wavelet and SVM [J]. Journal of Northeast Electric Power University, 2013, 34(2), 27—31.
- [6] 杨 茂,杨春霖. 基于模糊力计算的风电功率预测研究[J]. 东北电力大学学报, 2017, 37(5): 5—11.
- [6] Yang Mao, Yang Chunlin. Research on wind power real-time forecasting based on fuzzy granular computing [J]. Journal of Northeast Electric Power University, 2017, 37(5): 5—11.
- [7] Liu Hui, Tian Hongqi, Pan Difufu, et al. Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks[J]. Applied

- Energy, 2013, 107(6): 191—208.
- [8] 尹东阳, 盛义发, 蒋明洁, 等. 基于粗糙集理论-主成分分析的Elman神经网络短期风速预测[J]. 电力系统保护与控制, 2014, 42(11): 46—51.
- [8] Yin Dongyang, Sheng Yifa, Jiang Mingjie, et al. Short-term wind speed forecasting using Elman neural network based on rough set theory and principal components analysis[J]. Power System Protection and Control, 2014, 42(11): 46—51.
- [9] 杨锡运, 孙宝君, 张新房, 等. 基于相似数据的支持向量机短期风速预测仿真研究[J]. 中国电机工程学报, 2012, 32(4): 35—41.
- [9] Yang Xiyun, Sun Baojun, Zhang Xifang, et al. Short-term wind speed forecasting based on support vector machine with similar data[J]. Proceedings of the CSEE, 2012, 32(4): 35—41.
- [10] Pelikan E, Eben K, Resler J, et al. Wind power forecast by an empirical model using NWP output [A]. International Conference on Environment and Electrical Engineering[C], Prague, Czech Republic, IEEE, 2010.
- [11] 冬 雷, 王丽婕, 高 爽, 等. 基于混沌时间序列的大型风电场发电功率预测建模与研究[J]. 电工技术学报, 2011, 23(12): 125—128.
- [11] Dong Lei, Wang Lijie, Gao Shuang, et al. Modeling and analysis of prediction of wind power generation in the large wind farm based on chaotic time series [J]. Transactions of China Electrotechnical Society, 2011, 23(11): 125—128.
- [12] 范高锋, 王伟胜, 刘 纯, 等. 基于神经网络的风电功率预测[J]. 中国电机工程学报, 2008, 28(34): 118—123.
- [12] Fan Gaofeng, Wang Weisheng, Liu Chun, et al. Wind power prediction based on artificial neural networks [J]. Proceedings of the CSEE, 2008, 28(34): 118—123
- [13] 潘迪夫, 刘 辉, 李燕飞. 基于时间序列和卡尔曼滤波算法的风电场风速预测优化模型[J]. 电网技术, 2008, 32(7): 82—86.
- [13] Pan Difu, Liu Hui, Li Yanfei. A wind speed forecasting optimization model for wind farms based on times series analysis and Kalman filter algorithm [J]. Power System Technology, 2008, 32(7): 82—86.
- [14] 王志勇, 郭创新, 曹一家. 基于模糊粗糙集和神经网络的短期负荷预测方法[J]. 中国电机工程学报, 2005, 25(19): 7—11.
- [14] Wang Zhiyong, Guo Chuangxin, Cao Yijia. A method for load forecasting integrating fuzzy rough set with artificial neural network [J]. Proceedings of the CSEE, 2005, 25(19): 7—11
- [15] Li Gong, Shi Jing. On comparing three artificial neural networks for wind speed forecasting [J]. Applied Energy, 2010, 87(7): 2313—2320.
- [16] 丁志勇, 杨 萍, 杨 曦, 等. 基于连续时间段聚类的支持向量机风电功率预测方法[J]. 电力系统自动化, 2012, 36(14): 131—135.
- [16] Ding Zhiyong, Yang Ping, Yang Xi, et al. Wind power prediction method based on sequential time clustering support vector machine [J]. Automation of Electric Power Systems, 2012, 36(14): 131—135.
- [17] 龚 伟. 基于信息熵和互信息的流域水文模型不确定性分析[D]. 北京: 清华大学, 2012.
- [17] Gong Wei. Uncertainty analysis of watershed hydrological model based on information entropy and mutual information[D]. Beijing: Tsinghua University, 2012.
- [18] 马超群, 兰秋军, 陈为民. 金融数据挖掘[M]. 北京: 科学出版社, 2006.
- [18] Ma Chaoqun, Lan Qiuqun, Chen Weimin. Financial data mining [M]. Beijing: Science Press, 2006.
- [19] 雷杨娜, 孙 娴, 姜创业. 测风塔选址对复杂地形风电场风资源评估的影响[J]. 水电能源科学, 2013, 31(4): 236—239.
- [19] Lei Yangna, Sun Xian, Jiang Chuangye. Influence of wind tower selection on wind resource assessment in complex terrain wind farm [J]. Water resources and Power, 2013, 31(4): 236—239.
- [20] 杨 茂, 黄宾阳, 江 博, 等. 基于卡尔曼滤波和支持向量机的风电功率实时预测研究[J]. 东北电力大学学报, 2017, 37(2): 45—51.
- [20] Yang Mao, Huang Binyang, Jiang Bo, et al. Real-Time prediction for wind power based on Kalman filter and support vector machines [J]. Journal of Northeast Electric Power University, 2017, 37(2): 45—51.
- [21] 国家能源局. 风电场功率预测预报管理暂行办法[J]. 太阳能, 2011, (14): 6—7.
- [21] The National Energy Board. The notification of wind power forecasting management interim measures [J]. Solar Energy, 2011, (14): 6—7.

## RESEARCH ON REAL TIME PREDICTION OF WIND POWER BASED ON ENTROPY WEIGHT COMPREHENSIVE ASSOCIATION DATA MINING

Yang Mao<sup>1</sup>, Ma Jian<sup>1</sup>, Li Dayong<sup>2</sup>, Su Xin<sup>3</sup>, Sun Yong<sup>4</sup>

(1. *Modern Power System Simulation Control & Renewable Energy Tech. Jilin Province Key Laboratory(Northeast Electric Power University, Jilin 132012, China*; 2. *State Grid Jilin Electric Power Company Tonghua Power Supply Company, Tonghua 130022, China*;

3. *College of Science, Northeast Electric Power University, Jilin 132012, China*;

4. *State Grid Zibo Power Supply Company, Zibo 255000, China*)

**Abstract:** The correlation information of wind power history data can effectively improve the accuracy and computational efficiency of wind power prediction. This paper use a comprehensive correlation index based on entropy-weight to evaluate nonlinear mapping relationship between different historical periods of wind power sample and reference samples quantitatively, and solve the correlation redundancy between input and output variables of prediction model. And compare with the indicators of Pearson correlation coefficient, Kendall correlation coefficient, Spearman correlation coefficient and correlation coefficient of mutual information. Then, through intimate-samples selection, hidden layer structure optimization and network weights assignment, a modified model of real time wind power prediction is used to overcome the defect of the redundant degree training samples and slow convergence in traditional neural network training process, and improve the generalization ability and computational efficiency of the prediction model. Based on the measured data of wind farms in Jilin province, the simulation results show that the proposed method can effectively improve the accuracy of wind power forecasting.

**Keywords:** association information mining; entropy weight comprehensive correlation coefficient; correlation redundancy; model generalization ability