基于双维度顺序填补框架与改进 Kohonen 天气聚类的光伏发电短期预测

李燕青,杜莹莹
(华北电力大学 河北省输变电设备安全防御重点实验室,河北 保定 071003)

摘要:为提高部分数据缺失情况下智能电网光伏发电功率预测精度,提出一种基于双维度顺序填补框架与改进 Kohonen 天气聚类的光伏发电功率预测模型。采用双维度顺序填补方法补齐缺失数据,基于完整数据分析光伏发电功率影响因素,建立改进 Kohonen 天气聚类模型,并利用 S-Kohonen 实现预测日天气类型识别,将聚类历史日数据与预测日气象数据作为输入,采用多种群果蝇优化广义回归神经网络(MFOA-GRNN)模型对预测日光伏发电功率进行预测。仿真结果表明,所提方法能有效提高预测精度,为实现数据缺失情况下智能电网光伏发电功率的精准预测提供了一种思路。

关键词:智能电网;光伏功率预测;双维度顺序填补;改进 Kohonen;天气聚类
中图分类号:TM 615 **文献标识码**:A DOI:10.16081/j.issn.1006-6047.2019.01.009

0 引言

智能电网以发电、输电、变电、配电、用电及调度等多个环节,即覆盖电力系统整个生产过程的全景实时系统为目标[1-2]。其中,光伏发电作为一种将光能直接转化为电能的技术,具有可再生、无公害等显著优点[3-4]。随着光伏发电技术的发展,光伏并网过程的稳定性、可靠性成为智能电网需要关注的重点问题。光伏输出功率所具有的间歇性、波动性等特点对并网过程产生的不利影响给相关调度部门带来了极大的挑战[5-6]。为此,精准地预测光伏发电功率是实现光伏安全稳定并网必不可少的环节。

光伏发电功率预测根据不同的数据条件具有不同的预测模式,探寻已知变量数据与待预测光伏发电功率间的映射关系是实现这些预测模式的关键[7-8]。因此,保证已知变量数据的完整性、数据聚类的有效性显得尤为重要。现阶段,对于光伏发电预测数据缺失处理方法的研究较少。文献[9]针对天气类型信息缺失的问题,以辐照度特征参数建立支持向量机(SVM)模型,实现部分天气类型历史数据的补齐,但这种方法不适用于辐照度数据也部分缺失的情景。文献[10]提出采用 K 近邻补全法对缺失的历史数据进行补齐,但其仅以样本点间欧氏距离来确定近邻节点,数据填补精度有待提高。对于精细化历史数据聚类的研究,文献[11]分析了用于辐照度建模的确定性和随机性独立模型存在的不足,采用改进的模糊均值聚类(FCM)算法对辐照特性进行分时段聚类,经测试后验证了模型的有效性,但没有计及温度、湿度等关键性影响因素。文献[12]考虑天气类型对光伏发电功率的影响,采用智能水滴算法优化 Elman 神经网络作为预测模型,分别给出了晴天、阴天、雨天 3 种天气类型下预测日的光伏功率。文献[13]通过概率神经网络(PNN)算法划分晴、多云、阴、雨 4 种天气类型,对气象因素进行主成分分析(PCA)降维后,建立分散搜索-支持向量机回归(SS-SVR)预测模型。文献[14]利用辐照度、温度、风速 3 个变量的依赖性及其时间信息,将天气类型分为冬季和夏季,采用时空算法预测光伏发电功率。上述 3 种方法均考虑了天气类型对光伏发电功率的影响,并取得了良好的预测效果,但对于天气类型的划分比较粗略,没有达到精细化要求。

基于上述研究,为了提高缺失数据填补与天气聚类的精准性,进而提高光伏功率预测精度,本文提出一种基于双维度数据填补与改进 Kohonen 天气聚类的光伏功率预测方法。采用双维度数据填补框架对缺失的气象数据与光伏功率数据进行填补,利用完整历史数据集分析影响光伏发电功率的各个因素;选取适当因素用于 Kohonen 天气聚类,基于聚类结果对预测日进行 S-Kohonen 天气类型的划分;以同类型历史日数据与预测日气象数据为输入,建立多种群果蝇优化广义回归神经网络(MFOA-GRNN)预测模型,并对比基于全部历史日数据的预测结果,验证所提预测方法的精准性。

1 基于双维度顺序填补框架的数据补齐

历史气象数据及光伏功率数据在采集过程中因测量、传输、存储设备故障,可能会出现数据缺失的情况,为实现光伏出力的精准预测,需要先对缺失数据进行有效填补。在文献[15]的基础上,提出基于双维度顺序填补框架的数据补齐方法,考虑时间、属性双维度相似性,为更好地反映数据间的影响与相

收稿日期:2018-02-08;**修回日期**:2018-06-17
基金项目:自治区科技支疆研究项目(2017E0277)
Project supported by the Science and Technology Reconstruction-Assistant for Xinjiang Uygur Autonomous Region(2017E0277)

关程度,以灰色相关度代替欧氏距离进行缺失数据的补齐。数据填补框架充分利用数据特有的时间、属性相关性来衡量不同数据源间相似度,填补过程中先确定缺失数据的填补顺序后再对缺失数据进行填补,将已补齐的缺失数据作为观测值用于后续填补,这对缺失数据较密集时由缺失数据源的完整近邻与其相似度较低引起的填补精度下降问题起到缓解作用。

令各气象因素与光伏发电功率组成的 N 个数据源的集合为 $S=\{S_1,S_2,\cdots,S_i,\cdots,S_N\}$,其中数据源 $S_i=(x_{i1},x_{i2},\cdots,x_{iM})$ 包含 M 个时间点的数据。基于双维度顺序填补框架实现数据补齐的步骤如下。

a. 确定时间与属性相关性模型。采用灰色关联度来度量数据源间双维度的相似度,则数据源 S_i 和 S_j 之间的双维度相似度为:

$$S_{imij}=\frac{1}{1+\min\{\gamma_{ij}^t,\gamma_{ij}^a\}} \tag{1}$$

其中,γ_{ij}^t 为时间维度相似度;γ_{ij}^a 为属性维度相似度。设 T 时刻数据源 S_i 中待补缺失数据 x_{iT} 的时间邻域为前 t 个时间点的数据,即 $D'_i=(x'_{i1},x'_{i2},\cdots,x'_{it})$,属性邻域为与待补缺失数据具有相关性的因子数据,即 $D_i=(x_{i1},x_{i2},\cdots,x_{ia})$。对应于数据源 S_j 中的时间邻域和属性邻域分别为 $D'_j=(x'_{j1},x'_{j2},\cdots,x'_{jt})$ 和 $D_j=(x_{j1},x_{j2},\cdots,x_{ja})$,则:

$$\gamma_{ij}^t=\frac{1}{t}\sum_{k=1}^{t}\frac{\min\limits_j\min\limits_k|x'_{ik}-x'_{jk}|+\rho\max\limits_j\max\limits_k|x'_{ik}-x'_{jk}|}{|x'_{ik}-x'_{jk}|+\rho\max\limits_j\max\limits_k|x'_{ik}-x'_{jk}|} \tag{2}$$

$$\gamma_{ij}^a=\frac{1}{a}\sum_{k=1}^{a}\frac{\min\limits_j\min\limits_k|x_{ik}-x_{jk}|+\rho\max\limits_j\max\limits_k|x_{ik}-x_{jk}|}{|x_{ik}-x_{jk}|+\rho\max\limits_j\max\limits_k|x_{ik}-x_{jk}|} \tag{3}$$

其中,$\rho\in(0,1)$,为调整参数。

b. 筛选待补缺失数据源近邻节点。根据式(1)—(3)求取缺失数据源与剩余数据源间的双维度相似度。以剔除相似度小于给定双维度相似度阈值 σ 的节点为筛选规则。

c. 构建待补缺失数据源相似框架。由步骤 **b** 可得 T 时刻 n_q 个缺失数据源近邻节点的相似度列表。相似框架中,框架的顶点仅包含缺失数据源,以缺失数据源与其所有相对完整近邻节点间相似度融合后作为顶点权重,框架的边表示所连 2 个缺失数据源互为近邻节点,以二者的相似度作为边权重。所构建的相似框架直观反映了缺失数据源间的无向加权依赖关系。设 S_i 为缺失数据源,则权值计算公式为:

$$w_{S_i}=1-\prod_{S_k\in Q(S_i)}(1-S_{imik}) \tag{4}$$

其中,$Q(S_i)$ 为缺失数据源 S_i 的相对完整近邻节点集合;S_{imik} 为数据源 S_i 与 S_k 之间的双维度相似性。

d. 确定填补顺序。对于含 n_q 个缺失数据源的可能填补顺序 s_e,其对应的有向权相似框架可看作贝叶斯网络,其顶点权重是网络中每个顶点的先验概率,边权重是两顶点之间的条件概率,贝叶斯网络的联合概率是其对应填补顺序的置信度,计算公式为:

$$b(s_e)=\prod_{i=1}^{n_q}p(S_i\mid S_{Q(i)}) \tag{5}$$

其中,$S_{Q(i)}$ 为用于缺失数据源 S_i 填补的缺失近邻节点的集合,被填补后的缺失数据源可以被视为观测值并在后续填补过程中使用;$p(S_i\mid S_{Q(i)})$ 为用于填补的近邻节点列表加入 $S_{Q(i)}$ 后更新的顶点权重值,其计算公式如式(6)所示。

$$p(S_i\mid S_{Q(i)})=\begin{cases}w_{S_i} & S_{Q(i)}=\varnothing\\ 1-(1-w_{S_i})\prod\limits_{S_j\in S_{Q(i)}}(1-S_{imij}) & S_{Q(i)}\neq\varnothing\end{cases} \tag{6}$$

对于缺失数据源 S_i,$p(S_i\mid S_{Q(i)})$ 越大,填补误差越小;对于给定的一个填补顺序,当其置信度越高时,填补误差越小。因此,最优填补顺序问题可转化为选取最高置信度的贝叶斯网络问题,令 n_q 个缺失数据源所有可能的填补顺序集合 $S_E=\{s_{e1},s_{e2},s_{e3},\cdots\}$,填补顺序的确定转变为求 S_E 中置信度最大的序列,即:

$$s_e^*=\operatorname{argmax}(b(s_{ei}))\quad s_{ei}\in S_E \tag{7}$$

通过枚举所有可能的填补顺序,可以选择置信度最大的序列作为最优填补顺序,从而得到最优解。

e. 填补顺序确定后,采用基于近邻节点的缺失值填补算法(NI)进行缺失数据填补[15]。令待填补数据源 S_i 的近邻节点集合 $S_{Ni}=\{S_1,S_2,\cdots,S_{|S_{Ni}|}\}$,$S_i$ 的填补值 x'_i 为:

$$x'_i=\frac{\sum\limits_{S_j\in S_{Ni}}S_{imij}x_j}{\sum\limits_{S_j\in S_{Ni}}S_{imij}} \tag{8}$$

2　光伏发电影响因素分析

2.1　季节和天气类型对光伏发电的影响

本文以澳大利亚某光伏电站 2015 年 12 月至 2016 年 11 月 4 个季节 366 d 时间段 06:30—17:30 每 15 min 的气象数据和光伏发电功率数据作为研究对象。数据处理过程中,发现气象数据和光伏数据部分缺失,采用双维度顺序填补法对缺失数据填补后进行后续分析。图 1 为春、夏、秋、冬 4 个季节每天 45 个时间点的光伏出力情况。对比 4 个季节的光伏出力情况可知,春、夏、秋 3 个季节光伏发电功率没有太大差异,冬季光伏发电功率相较于春、夏、秋 3 个季节明显更小,说明在光伏功率预测问题

中需要考虑季节对光伏发电的影响,分春夏秋和冬季2个时间段进行光伏功率预测。观察4个季节每天的光伏发电功率可知,同一个季节不同日的光伏发电功率存在较大差异,需要考虑日天气类型对光伏发电功率的影响,可通过天气聚类提高预测模型精度。

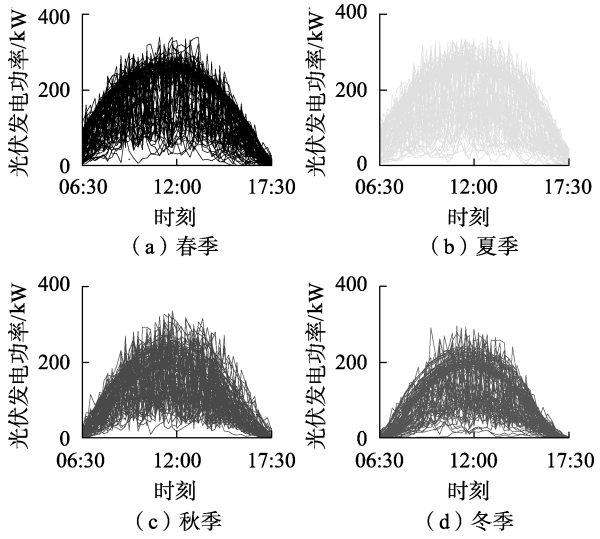


图1 4个季节每日光伏出力

Fig.1 Daily photovoltaic power of four seasons

2.2 各气象因素对光伏发电的影响

光伏电站提供的气象数据包含温度、湿度、辐照度、风速、风向。为消除各变量量纲不同产生的影响,将气象数据和光伏出力数据进行归一化处理。取某一周的气象及光伏出力数据绘制出其归一化后的曲线,如图2所示。可见光伏出力和温度、湿度、辐照度的变化趋势一致,且对辐照度的跟随趋势更加明显,而风速、风向和光伏出力之间没有显著的相关关系。故在光伏发电功率预测模型中,将温度、湿度、辐照度作为影响因子。

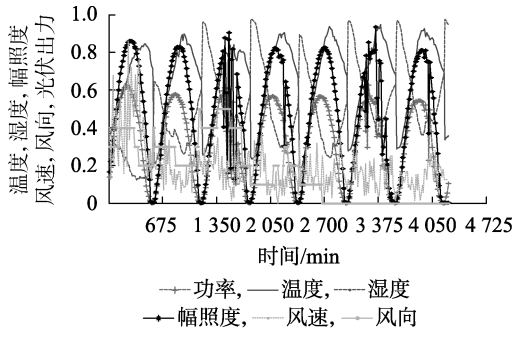


图2 光伏出力与各气象因素的归一化曲线

Fig.2 Normalized curves of photovoltaic power and meteorological factors

3 预测模型

3.1 改进 Kohonen 天气聚类

Kohonen网络是芬兰Teuvo Kohonen教授提出的一种无监督学习网络,属于自组织竞争型神经网络范畴[16]。Kohonen网络输入层与输出层连接初始权值的设定对聚类效果会产生一定的影响,采用给输出层每个神经元加上一个阈值 $\varepsilon_{\text{thre}}$ 的方法对网络进行优化[17]。采用改进的Kohonen神经网络进行训练的步骤如下。

a. Kohonen网络初始化。初始化网络权值 w_{ij} ($i=1,2,\cdots,n$;$j=1,2,\cdots,m$),$\boldsymbol{w}$ 为区间[0,1]内 $n\times m$ 阶的随机数矩阵。w_{ij} 满足的约束条件为:

$$\sum_{i=1}^{n} w_{ij}=1 \tag{9}$$

b. 计算输入样本 $\boldsymbol{X}$ 与竞争层神经元间距离 d_j 为:

$$d_j=\left|\sum_{i=1}^{n}(x_i-w_{ij})^2\right|+\varepsilon_{\text{thre}}(b) \quad j=1,2,\cdots,m \tag{10}$$

其中,$\varepsilon_{\text{thre}}(b)$ 为区间[0,1]内的随机数。

c. 选择神经元。将满足输入样本与竞争层神经元间距离为最小值时的输出层神经元 b' 作为最优匹配输出神经元。

d. 修正神经元 b' 和在其邻域 $N_c(t')$ 内包含的节点权值,其邻域的范围是与神经元 b' 间的欧氏距离小于邻域半径的所有节点 t'。

$$N_c(t')=\{t'\mid \text{norm}(p_{t'},p_{b'})<r\} \quad t'=1,2,\cdots,n \tag{11}$$

$$w_{ij}(k+1)=w_{ij}(k)+\eta(x_i-w_{ij}) \tag{12}$$

其中,norm(·)表示计算两神经元间欧氏距离;η 为学习速率。

e. 训练阈值。$\varepsilon_{\text{thre}}$ 的训练规则为:若某神经元的获胜率高,则其 $\varepsilon_{\text{thre}}$ 逐渐变小;若某神经元的获胜率低,则其 $\varepsilon_{\text{thre}}$ 逐渐变大。

f. 返回步骤**b**,直到连接权值调整量很小为止。

完成上述训练过程,可实现天气类型的聚类。

3.2 S-Kohonen 识别待预测日类型

通过在竞争层之后增加输出层可以使Kohonen网络成为有监督学习的S-Kohonen网络,输出层的每个节点表示一类数据,改进S-Kohonen网络结构如图3所示。因此,在修正权值时,还需要修正竞争层优胜节点领域内节点和输出层节点之间的权值[18],修正方式为:

$$w_{jh}(k+1)=w_{jh}(k)+\eta'(Y_h-w_{jh}) \tag{13}$$

其中,η' 为学习概率;Y_h 为样本所属类型。

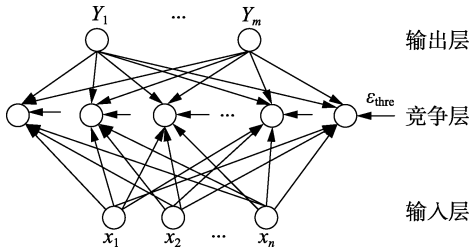


图3 改进S-Kohonen网络结构

Fig.3 Improved S-Kohonen network structure

完成网络训练后,可实现对待预测日天气类型的分类。

3.3　基于 MFOA-GRNN 的光伏功率预测模型

广义回归神经网络(GRNN)是美国学者于 1991 年提出的一种径向基神经网络,由输入层、模式层、求和层和输出层构成[19-20]。令网络输入 $\boldsymbol{X}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$,对应的输出 $\boldsymbol{Y}=[y_1,y_2,\cdots,y_m]^{\mathrm{T}}$。则模式层神经元传递函数为:

$$p_i=\exp\left[-\frac{(\boldsymbol{X}-\boldsymbol{X}_i)^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{X}_i)}{2\sigma^2}\right]\quad i=1,2,\cdots,n \tag{14}$$

其中,$\boldsymbol{X}_i$ 为第 i 个神经元对应的学习样本;σ 为高斯函数的宽度系数,称为光滑因子。

求和层 2 种类型的传递函数分别为:

$$S_{\mathrm{D}}=\sum_{i=1}^{n}p_i \tag{15}$$

$$S_{\mathrm{N}j}=\sum_{i=1}^{n}y_{ij}p_i\quad j=1,2,\cdots,m \tag{16}$$

其中,y_{ij}为模式层第 i 个神经元与求和层第 j 个分子求和神经元间的连接权值。

网络输出层函数为:

$$y_j=\frac{S_{\mathrm{N}j}}{S_{\mathrm{D}}}\quad j=1,2,\cdots,m \tag{17}$$

由上式可知,σ 的取值会对预测效果产生很大的影响,因此采用 MFOA-GRNN 确定最佳光滑因子,建立改进 MFOA-GRNN 模型的具体步骤见文献[21]。

4　算例分析

4.1　光伏发电功率预测

以第 2 节中的气象数据和光伏发电功率数据作为研究对象,其中气象数据和光伏数据部分缺失,且存在部分缺失数据较为密集的情况。采用双维度顺序填补法对缺失数据填补,枚举所有可能的填补顺序后选择置信度最大的序列作为最优填补顺序得到最优填补结果,其中部分填补顺序下的置信度(置信度较高的部分)如表 1 所示。由表 1 可知,按温度-功率-湿度-辐照度-风速-风向的填补顺序可以得到最优填补结果,将得到的完整数据集用于后续建模分析。

表 1　部分填补顺序下的置信度

Table 1　Confidence levels under different imputation orders

填补顺序	置信度
温度-湿度-功率-辐照度-风向-风速	0.686 4
功率-辐照度-温度-湿度-风速-风向	0.691 3
温度-辐照度-功率-湿度-风速-风向	0.718 2
功率-温度-湿度-辐照度-风速-风向	0.746 8
温度-功率-湿度-辐照度-风速-风向	0.781 2

取每日最高温度 $T_{\max}$、平均温度 T_{mean}、最高湿度 $H_{\max}$、平均湿度 H_{mean}、最大辐照度 $W_{\max}$作为改进 Kohonen 神经网络的输入特征向量,分冬季(6—8 月)和春夏秋 2 个时间段进行天气聚类,设置 2015 年 12 月 1 日的日序号为 1,改进 Kohonen 神经网络输入层节点数为 5,竞争层节点排列在 7×7 阶的方阵中且节点数为 49,最大和最小邻域分别为 1.5 和 0.5,最大和最小学习概率分别为 0.1 和 0.01,网络学习调整次数为 10 000 次。冬季时间段,以 6 月 1 日至 8 月 31 日的数据作为样本,预测 8 月 1 日至 8 月 31 日光伏发电功率,对 8 月 31 日预测结果进行定性分析,其历史日天气聚类情况如附录中表 A1 所示;春夏秋时间段,预测 10 月 1 日至 10 月 31 日光伏发电功率,对 10 月 1 日预测结果进行定性分析,其历史日天气聚类情况如附录中表 A2 所示。由表 A1 和表 A2 可知,冬季时间段和春夏秋时间段历史日天气聚类后分别被分为 4 种和 6 种天气类型,日序号相近的历史日聚为一类的可能性较大,但也存在日序号相差较大的历史日数据聚为一类的情况。

基于上述聚类结果,采用 S-Kohonen 神经网络进行待预测日天气类型识别,分别设置冬季和春夏秋时间段的输出层节点数为 4 和 6。得到冬季时间段预测日天气类型属于第 4 类,春夏秋时间段预测日天气类型属于第 3 类。冬季时间段,以表 A1 中第 4 类 32 d 的温度、湿度、辐照度、光伏发电功率及预测日温度、湿度、辐照度数据作为输入,预测日光伏发电功率作为输出,建立基于 MFOA-GRNN 的光伏功率预测模型。春夏秋时间段,以表 A2 中第 3 类历史日数据作为输入。图 4、5 分别为冬季和春夏秋时间段预测日 06:30—17:30 每隔 15 min 的光伏发电

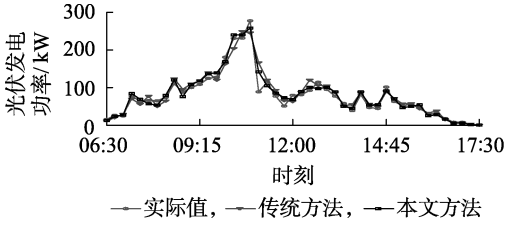


图 4　冬季时间段预测日光伏发电功率预测结果

Fig.4　Forecasting results of photovoltaic power in winter time period

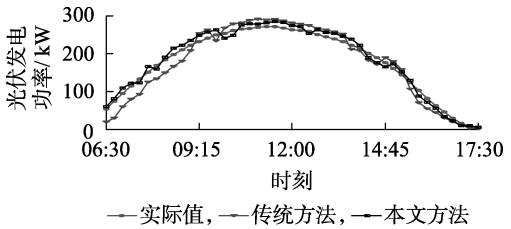


图 5　春夏秋时间段预测日光伏发电功率预测结果

Fig.5　Forecasting results of photovoltaic power in spring, summer and autumn time period

功率实际值、基于所有历史日的 MFOA-GRNN 预测值(简称传统方法)及基于聚类历史日的 MFOA-GRNN 预测值(简称本文方法)。

4.2 结果分析

从图 4、5 可以看出,基于所有历史日的 MFOA-GRNN 预测与基于聚类历史日的 MFOA-GRNN 预测的光伏出力曲线变化趋势均与实际一致。为更好地评估对比预测效果,采用均方根误差(RMSE)与平均绝对百分比误差(MAPE)对冬季和春夏秋 2 个时间段模型预测性能进行评价。表 2 为 2 个时间段的定性分析日预测结果评价,表 3 为 2 个时间段的多日预测结果评价。由表 2、3 可知,基于聚类历史日的 MFOA-GRNN 模型预测效果更好,这说明双维度顺序填补法保证了预测准确性,Kohonen 天气聚类提高了模型预测精度。通过比较发现,春夏秋时间段聚类前、后模型预测效果相差较大,可能是因为春、夏、秋 3 个季节历史日数量比冬季时间段多,天气类型间的差异相对较大,聚类对模型的预测精度影响更大。

表 2 单日预测结果比较

Table 2 Comparison of forecasting results for single-day

预测方法	时间	RMSE/kW	MAPE/%
传统方法	8 月 31 日	18.41	19.55
	10 月 1 日	22.35	19.44
本文方法	8 月 31 日	12.36	14.17
	10 月 1 日	11.45	11.68

表 3 多日预测结果比较

Table 3 Comparison of forecasting results for multi-days

预测方法	时间段	RMSE/kW	MAPE/%
传统方法	8 月 1—31 日	18.72	17.65
	10 月 1—31 日	20.14	20.56
本文方法	8 月 1—31 日	13.19	14.55
	10 月 1—31 日	12.23	13.18

5 结论

本文在智能电网的背景下,考虑历史数据缺失问题,提出双维度顺序填补框架,按温度-功率-湿度-辐照度-风速-风向的填补顺序进行最优数据填补。为提高天气类型划分的精细性,分析光伏发电功率影响因素后,选取温度、湿度、辐照度作为影响因子,分春夏秋和冬季 2 个时间段基于完整历史数据采用改进 Kohonen 神经网络聚类,分别得到 4 种和 6 种天气类型,并利用 S-Kohonen 神经网络识别 2 个时间段的预测日天气类型。采用 MFOA-GRNN 模型进行预测日光伏功率预测,对比基于聚类历史日与基于全部历史日的预测情况可知,单日与多日预测中基于聚类历史日的模型预测效果都要更好,说明双维度顺序填补框架保证了填补数据的有效性,改进 Kohonen 神经网络天气聚类方法提高了天气聚类的精准性。因此,基于双维度顺序填补框架与改进 Kohonen 天气聚类的方法能提高光伏功率预测精度,为历史数据部分缺失情景下的智能电网光伏发电功率预测提供一种有效策略。

本文只考虑了历史数据缺失的情景,下一步的研究将关注历史坏数据的剔除与修正。

附录见本刊网络版(http://www.epae.cn)。

参考文献:

[1] 鞠平,周孝信,陈维江,等. "智能电网+"研究综述[J]. 电力自动化设备,2018,38(5):2-11.
JU Ping,ZHOU Xiaoxin,CHEN Weijiang,et al. "Smart grid plus" research overview [J]. Electric Power Automation Equipment,2018,38(5):2-11.

[2] 孙丰杰,王承民,谢宁. 面向智能电网大数据关联规则挖掘的频繁模式网络模型[J]. 电力自动化设备,2018,38(5):110-116.
SUN Fengjie,WANG Chengmin,XIE Ning. Frequent pattern network model for association rule mining of big data in smart grid[J]. Electric Power Automation Equipment,2018,38(5):110-116.

[3] HU C,TAN B,YAO X. Research voltage stability of photovoltaic power generation system[C]// 2016 IEEE Electronic and Automation Control Conference(IMCEC). Xi'an,China:IEEE,2016:1491-1494.

[4] 贾祺,严干贵,李泳霖,等. 新型多端输入光伏并网系统运行控制策略[J]. 电力自动化设备,2017,37(10):43-48.
JIA Qi,YAN Gangui,LI Yonglin,et al. Operation control strategy of novel multi-input grid-connected PV system[J]. Electric Power Automation Equipment,2017,37(10):43-48.

[5] 李升,卫志农,孙国强,等. 大规模光伏发电并网系统电压稳定分岔研究[J]. 电力自动化设备,2016,36(1):17-23.
LI Sheng,WEI Zhinong,SUN Guoqiang,et al. Voltage stability bifurcation of large-scale grid-connected PV system [J]. Electric Power Automation Equipment,2016,36(1):17-23.

[6] 王继东,庞文杰. Fisher 判别分类法在光伏并网系统稳态电能质量评估中的应用[J]. 电力自动化设备,2017,37(3):50-54.
WANG Jidong,PANG Wenjie. Application of Fisher discriminant analysis in steady-state power quality evaluation of grid-connected photovoltaic system [J]. Electric Power Automation Equipment,2017,37(3):50-54.

[7] 龚莺飞,鲁宗相,乔颖,等. 光伏功率预测技术[J]. 电力系统自动化,2016,40(4):140-151.
GONG Yingfei,LU Zongxiang,QIAO Ying,et al. An overview of photovoltaic energy system output forecasting technology[J]. Automation of Electric Power Systems,2016,40(4):140-151.

[8] 黎静华,赖昌伟. 考虑气象因素的短期光伏出力预测的奇异谱分析方法[J]. 电力自动化设备,2018,38(5):50-57,76.
LI Jinghua,LAI Changwei. Singular spectrum analysis method for short-term photovoltaic output prediction considering meteorological factors[J]. Electric Power Automation Equipment,2018,38(5):50-57,76.

[9] 王飞,米增强,甄钊,等. 基于天气状态模式识别的光伏电站发电功率分类预测方法[J]. 中国电机工程学报,2013,33(34):75-82.
WANG Fei,MI Zengqiang,ZHEN Zhao,et al. A classified forecasting approach of power generation for photovoltaic plants based on weather condition pattern recognition[J]. Proceedings of the CSEE,2013,33

(34):75-82.

[10] 史佳琪,张建华. 基于深度学习的超短期光伏精细化预测模型研究[J]. 电力建设,2017,38(6):28-35.
SHI Jiaqi,ZHANG Jianhua. Ultra-short-term photovoltaic refined forecasting model based on deep learning[J]. Electric Power Construction,2017,38(6):28-35.

[11] 郭一飞,高厚磊,田佳. 引入聚类分析的光伏出力建模及其在可靠性评估中的应用[J]. 电力系统自动化,2016,40(23):93-100.
GUO Yifei,GAO Houlei,TIAN Jia. Photovoltaic output modeling by introducing clustering analysis and its application in reliability evaluation[J]. Automation of Electric Power Systems,2016,40(23):93-100.

[12] 赵文清,郭丙旭,李刚,等. 基于智能水滴算法优化 Elman 神经网络的光伏电站输出功率预测[J]. 太阳能学报,2017,38(6):1553-1559.
ZHAO Wenqing,GUO Bingxu,LI Gang,et al. Output power forecast of plant based on Elman neural network drop algorithm[J]. Acta Energiae Solaris Sinica,2017,38(6):1553-1559.

[13] 王昕,黄柯,郑益慧,等. 基于 PNN/PCA/SS-SVR 的光伏发电功率短期预测方法[J]. 电力系统自动化,2016,40(17):156-162.
WANG Qi,HUANG Ke,ZHENG Yihui,et al. Short-term forecasting method of photovoltaic output power based on PNN/PCA/SS-SVR [J]. Automation of Electric Power Systems,2016,40(17):156-162.

[14] TASCIKARAOGLU A,SANANDAJI B M,CHICCO G,et al. A short-term spatio-temporal approach for photovoltaic power forecasting[C]// 2016 IEEE Power Systems Computation Conference(PSCC). Genoa, Italy:IEEE,2016:1-7.

[15] 马茜,谷峪,李芳芳,等. 顺序敏感的多源感知数据填补技术[J]. 软件学报,2016,27(9):2332-2347.
MA Qian,GU Yu,LI Fangfang,et al. Order-sensitive missing value imputation technology for multi-source sensory data[J]. Journal of Software,2016,27(9):2332-2347.

[16] 马小慧,阳育德,龚利武. 基于 Kohonen 聚类和 SVM 组合算法的电网日最大负荷预测[J]. 电网与清洁能源,2014,30(2):7-11.
MA Xiaohui, YANG Yude, GONG Liwu. Forecasting of the daily maximum load based on a combined model of Kohonen custering and SVM[J]. Power System and Clean Energy,2014,30(2):7-11.

[17] 方浩,王艳红. 改进的 Kohonen 神经网络航迹关联算法[J]. 计算机应用,2013,33(5):1476-1480.
FANG Hao,WANG Yanhong. Track correlation algorithm based on modified Kohonen neural network[J]. Journal of Computer Applications,2013,33(5):1476-1480.

[18] 郭欣,王蕾,宣伯凯,等. 基于有监督 Kohonen 神经网络的步态识别[J]. 自动化学报,2017,43(3):430-438.
GUO Xin,WANG Lei,XUAN Bokai,et al. Gait recognition based on supervised Kohonen neural network [J]. Acta Automatica Sinica,2017,43(3):430-438.

[19] 姚李孝,刘学琴,伍利,等. 基于广义回归神经网络的电力系统中长期负荷预测[J]. 电力自动化设备,2007,27(8):26-29.
YAO Lixiao,LIU Xueqin,WU Li,et al. Mid- & long-term load forecast based on GRNN[J]. Electric Power Automation Equipment, 2007,27(8):26-29.

[20] ANDREEV P G,YURKOV N K,ANDREEVA T V. Artificial neural networks of GRNN type in prediction problems[C]// 2017 IEEE Soft Computing and Measurements (SCM). Petersburg, Russia: IEEE, 2017:458-460.

[21] 李冬辉,尹海燕,郑博文. 基于 MFOA-GRNN 模型的年电力负荷预测[J]. 电网技术,2018,42(2):585-590.
LI Donghui,YIN Haiyan,ZHENG Bowen. An annual load forecasting model based on generalized regression neural network with multi-swarm fruit fly optimization algorithm[J]. Power System Technology, 2018,42(2):585-590.

作者简介:


李燕青

李燕青(1974—),男,河北石家庄人,教授,博士,主要研究方向为智能电网、新能源发电与微电网控制(**E-mail**:hdlyq@163.com);

杜莹莹(1994—),女,四川成都人,硕士研究生,主要研究方向为新能源发电与微电网控制(**E-mail**:duyingying2016@126.com)。

Short-term photovoltaic power forecasting based on double-dimensional sequential imputation framework and improved Kohonen clustering

LI Yanqing,DU Yingying

(Hebei Provincial Key Laboratory of Power Transmission Equipment Security Defense, North China Electric Power University,Baoding 071003,China)

Abstract:To improve the accuracy of smart grid photovoltaic power forecasting in the absence of partial data,a forecasting method based on double-dimensional sequential imputation framework and improved Kohonen clustering is proposed. Firstly,using double-dimensional sequential imputation method to complement missing data,analyzing the influential factors of photovoltaic power based on the complete data,an improved Kohonen weather clustering model is established. Secondly,the weather type of the forecasting day is identified using S-Kohonen method. Finally,the MFOA-GRNN(Multi-swarm Fruit fly Optimization Algorithm and Generalized Regression Neural Network) model is used to forecast the daily photovoltaic power using the meteorological data of clustering historical day and forecasting day as input. The results show that the proposed method can effectively improve the forecasting accuracy and provide a way to accurately forecast the smart grid photovoltaic power under the condition of missing data.

Key words:smart grid;photovoltaic power forecasting;double-dimensional sequential imputation;improved Kohonen;weather clustering

附录

表 A1　冬季历史日天气聚类结果

Table A1　Historical weather clustering results of winter

聚类类别	日序号
1	184、186、189、192、195、196、201、202、206、218、224、225、227、228、229、230、235、246、262、268
2	185、188、191、199、205、211、212、214、215、216、220、221、222、223、232、234
3	190、193、198、203、207、208、209、210、219、226、233、236、237、238、242、243、244、245、247、252、253、254、255
4	187、194、197、200、204、213、217、231、239、240、241、248、249、250、251、256、257、258、259、260、261、、263、264、265、266、267、269、270、271、272、273、274

表 A2　春夏秋季历史日天气聚类结果

Table A2　Historical weather clustering results of three seasons

聚类类别	日序号
1	9、25、35、36、47、51、52、53、54、59、66、67、118、136、143、144、148、151、153、154、156、158、160、162、164、165、167、168、169、170、172、173、175、176、177、178、179、180、181、182、284、290、293、296、304、328
2	1、2、8、15、23、31、43、57、69、75、78、88、89、100、114、117、120、122、124、125、129、130、131、132、133、134、137、138、139、140、141、142、145、147、149、150、152、155、157、159、161、163、166、174、183、279、281、286、287、295、298、299、301、303、307、311、312、313、315、324、336、337、347、356、357、360、365、366
3	4、5、21、29、30、37、44、45、50、64、79、87、102、110、111、112、115、116、121、123、128、135、146、171、288、294、305、309、310、321、323、338、340、341、342、344、345、348、350、352
4	17、22、28、38、40、56、80、104、109、126、127、276、277、278、282、285、308、314、316、325、327、329、330、332、333、335、339、349、354、

	355、359、363、364
5	3、6、13、14、18、19、20、39、42、55、58、60、76、77、82、83、84、86、94、95、97、98、105、106、119、283、291、292、297、300、302、318、331、334、343、346、353、358、361、362
6	7、10、11、12、16、24、26、27、32、33、34、41、46、48、49、61、62、63、65、68、70、71、72、73、74、81、85、90、91、92、93、96、99、101、103、107、108、113、280、289、317、319、320、322、326、351