

# 基于风电 SCADA 数据定子温度的预处理算法<sup>①</sup>

黄昊<sup>1,2</sup>, 易灵芝<sup>1</sup>, 詹俊<sup>2</sup>

<sup>1</sup>(湘潭大学, 湘潭 411105)

<sup>2</sup>(湖南优利泰克自动化系统有限公司, 长沙 410205)

**摘要:** 针对风电机组各部件性能分析过程繁琐低效、预测精度不高以及经济效益不足的非正常风电机组状态问题, 提出一种基于风电 SCADA 数据定子温度的预处理算法. 通过分析风电 SCADA 系统采集的各部件数据, 对于其中发电机的定子温度, 优化数据处理与分析的过程, 改进现有的最优组内方差预处理算法, 监测定子温度的趋势与非正常温度的状态, 提高了对发电机定子的维修效率. 通过实例分析表明改进后的最优组方差算法可行且高效, 能够准确处理发电机定子温度曲线数据并通过使用神经网络进行预测, 显著提高了风电机组中发电机定子温度预测的准确性.

**关键词:** 发电机定子温度; 数据预处理; 改进 OIV 算法; SCADA 系统

引用格式: 黄昊, 易灵芝, 詹俊. 基于风电 SCADA 数据定子温度的预处理算法. 计算机系统应用, 2018, 27(7): 193-198. <http://www.c-s-a.org.cn/1003-3254/6464.html>

## Preprocessing Algorithm of Stator Temperature Based on Wind Farm SCADA Data

HUANG Hao<sup>1,2</sup>, YI Ling-Zhi<sup>1</sup>, ZHAN Jun<sup>2</sup>

<sup>1</sup>(Xiangtan University, Xiangtan 411105, China)

<sup>2</sup>(Hunan UliTech Automation System Co. Ltd., Changsha 410205, China)

**Abstract:** A preprocessing algorithm for stator temperature based on wind power SCADA data has been put forward in view of deviant status of wind turbine, such as insufficiency on analytical process of performance of wind turbine, the predicting inaccuracy, and the deficiency of economic benefit. The maintaining efficiency of generator stator has been improved since the analysis on data gathered from each part by SCADA system of wind power. For the temperature of stator within the generator, the process and analysis of data have been optimized, with the amelioration of Optimal Interclass Variance (OIV) algorithm and the successful monitoring the trends of temperature of stator and its abnormal temperature status. The improved optimal interclass variance algorithm has been proved feasible and efficient, which is capable of processing date of generator stator's temperature curve data and making predictions via neural networks, while improving predicting accuracy of generator stator temperature significantly.

**Key words:** generator stator temperature; data preprocessing; improved Optimal Interclass Variance (OIV) algorithm; SCADA system

风力发电是目前技术最成熟、最具规模化开发条件的新能源发电方式之一, 因其清洁、高效、可再生等优点, 受到世界各国的重视. 风电机组是制约风电发

展的主要因素之一, 其良好的发电性能是保证风力发电企业生产运行管理和发展战略规划的决定因素<sup>[1]</sup>.

在风电迅猛发展的同时, 风机高额的运行维护成

① 基金项目: 国家自然科学基金 (61572416)

Foundation item: National Natural Science Foundation of China (61572416)

收稿时间: 2017-11-22; 修改时间: 2017-12-15; 采用时间: 2018-01-08; csa 在线出版时间: 2018-06-27

本影响了业内的经济效益。风场一般地处偏远、环境恶劣,并且机舱位于50~80 m以上的高空,给机组的维护维修工作造成了困难,增加了机组的运行维护成本。对于20年工作寿命的机组,运行维护成本估计占整个风场收入的10%~15%;对于海上风场,用于风机运行维护的成本高达风场收入的20%~25%<sup>[2,3]</sup>。高额的运行维护费用增加了业内的运营成本,大大降低了风力发电的实际经济效益。

目前很多企业通常使用SCADA系统采集多台风电机组各部件的状态、属性数据,不同机型记录的信号范围和类型可以相差很大,这样的SCADA系统融入到整个风电场中为降低损失、提高效益提供帮助<sup>[4]</sup>。考虑到SCADA系统实际采集到的数据复杂化的特点,传统数据预处理的清洗方法普遍缺乏通用性,且当存在多输入参数多维度或者缺少历史训练数据时,大部分用于数据预处理的清洗方法失效<sup>[5,6]</sup>。定子温度的实测曲线分布杂乱无序,无法直接用于风机部件的状态分析。在国内外风电领域中,应用较为广泛的聚类、统计分析等数据方法,但其清洗过程较为繁琐且条件苛刻,且仍需要针对风机机组具体的分部件进行数据处理,会显得力不从心,加上风电机组所处的环境导致各种因素的变化无常,SCADA系统所采集的相关数据并不能直接用来进行曲线分析,需要一种有效且通用性强的预处理算法对其进行清洗。

发电机是风力发电机组的核心部件,负责将经由风速导致的旋转的机械能转化为电能,并为电气系统供电。发电机长期运行于环境变化不定和电磁环绕中,容易发生故障。常见的故障有发电机振动过大、发电机过热、轴承过热、转子/定子线圈短路、定子温度过高等。据统计,在发电机所有故障中,轴承故障率为40%,定子的故障率为38%,转子的故障率为10%,其他故障占12%<sup>[7,8]</sup>。

只有经历过清洗预处理之后的曲线方能准确地反应发电机定子温度的状态,因此在风电机组性能分析中,对原始的SCADA数据进行清洗预处理极其重要。固为了提高对发电机定子的温度状态研究的可靠性,本文提出了一种用于曲线分析的改进最优组方差算法。该算法只需要对发电机定子温度和定子温度的相关参数进行分析,不仅仅改变了传统算法对多维度参数的依赖性,还能够准确的发现定子温度的非正常点,为发电机的维护提供了准确的数据支持。

## 1 数据预处理算法的研究

### 1.1 SCADA系统采集的定子温度相关参数

本文主要针对部件为发电机定子的温度,必然需要定子温度的相关参数和定子温度形成曲线进行分析。据了解<sup>[9]</sup>,发电机定子温度与电流、线圈绕组及铜耗有关:

$$P_{\text{铜}} = \omega IR^2 \quad (1)$$

其中, $\omega$ 为该发电机的相数(单相或三相), $I$ 为发电机电流, $R$ 为定子电阻,发热量 $Q \approx P$ 。

根据式(1)得知,发电机电流 $I$ 是SCADA系统能够采集的数据,而 $R$ 即绕组在本文默认为固定值, $\omega$ 为三相。由此得知,铜耗 $P_{\text{铜}}$ 与发电机电流密切相关,而定子温度与铜耗保持绝对的关系,即发电机定子温度的状态与发电机电流 $I$ 直接相关。

而作为定子温度的主要相关因素的发电机电流,将和定子温度形成的数据集通过预处理算法进行数据清洗。只需分别将定子温度与发电机电流等相关参数通过改进的最优组方差算法<sup>[10]</sup>和神经网络<sup>[11-13]</sup>工具箱进行分析,不仅改变了传统分析方法对繁杂数据的依赖性,而且通过现有SCADA采集的数据能够准确识别出发电机定子非正常温度,为发电机定子温度的状态监测提供可靠的数据支持。实例分析表明,使用该方法能够对定子温度的相关参数进行清洗预处理,具有很好的通用性,有助于实现风力发电机定子温度的状态监测及后期的准确诊断与评估,同时增强风电机组运行稳定性、降低维护成本、提高发电量和经济效益,均有重要意义。

### 1.2 原最优组内方差算法

已知定子温度曲线样本集合 $U$ ,设 $V = \{V_n, V_1\}$ 为目标集合,其中 $V_n$ 表示该集合中定子温度正常的数据集, $V_1$ 表示定子温度偏高或偏低的数据集(定义为非正常),且 $V$ 满足 $V_n \cap V_1 = \emptyset$ , $V_n \cup V_1 = U$ 。

最优组内方差(Optimal Interclass Variance, OIV)算法要解决的问题就是如何快速、准确地对数据集 $U$ 进行清洗,识别出发电机定子温度正常和非正常的数据集,再进一步保留正常、剔除非正常,这实际上是个数据预处理的问题。

定义1.已知样本集合 $U = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ,且满足 $y_i > y_{i-1}, i \in (2, n)$ , $x$ 表示输入参数, $y$ 表示定子温度, $n$ 为数据点总数。有方差 $\varphi$ 公式:

$$\varphi = \frac{1}{\gamma} (y_j - \bar{y}_\gamma)^2 \tag{2}$$

原最优组方差 (OIV) 算法公式为:

$$\gamma = \arg \max \left\{ \sum_{\gamma=2}^n \left( \frac{1}{\gamma} \sum_{j=1}^{\gamma} (y_j - \bar{y}_\gamma)^2 \right) < S \right\} \tag{3}$$

$$\bar{\delta} = \sum_{\gamma=2}^n \left( \frac{1}{\gamma} \sum_{j=1}^{\gamma} (y_j - \bar{y}_\gamma)^2 \right) = \sum_{j=1}^{\gamma} \sum_{j=1}^{\gamma} \varphi \tag{4}$$

其中,  $y_j$  为  $j$  个点的定子温度;  $\bar{y}_\gamma$  为第 1 到第  $\gamma$  个点的定子温度平均数,  $\gamma$  为常数;  $\bar{\delta}$  为定子温度平均滑差值,  $S$  为方差阈值.

令发电机电流区间  $t=0, 1, 2, \dots, W$ , 其中  $W$  是发电机电流区间总数, 电流区间间隔  $T$  默认取 5A; 在某个电流区间内, 当且仅当满足式 (3) 时认为定子温度正常, 则有:

$$U_n(t) = \left\{ (x_i, y_i) \mid \begin{matrix} (x_i, y_i) \in U(t), \\ 1 \leq i \leq \gamma \end{matrix} \right\} \tag{5}$$

$$U_1(t) = \left\{ (x_i, y_i) \mid \begin{matrix} (x_i, y_i) \in U(t) \\ (x_i, y_i) \notin U_n(t) \end{matrix} \right\} \tag{6}$$

式中,  $U_n(t)$  是第  $t$  个发电机电流区间的正常数据集;  $U_1(t)$  是第  $t$  个发电机电流区间的非正常数据集.

最后, 对每个电流区间的数据处理结果进行归类整理, 得到发电机定子总的正常数据集和非正常数据集:

$$V_n = \{U_n(1), U_n(2), \dots, U_n(W)\} \tag{7}$$

$$V_1 = \{U_1(1), U_1(2), \dots, U_1(W)\} \tag{8}$$

### 1.3 OIV 算法的使用分析

输出曲线反应了定子的状态, 为形象说明 OIV 算法的思想和求解特点, 下面以一台风电机组中发电机定子的实测温度曲线单月份的数据集为例来进一步说明. 发电机定子实测温度曲线分布如图 1 所示.

取电流区间间隔  $T=5$  A, 暂时实验以方差阈值  $S < \bar{\delta}$ , 根据电流区间可将原始数据集划分为 37 个状态量, 有  $U = \{U(0), U(2), \dots, U(36)\}$ . 由于每个状态量的求解过程相似, 这里仅以 10 号状态量  $U(10)$  为例进行详细说明, 同时结合 5 号、20 号区间绘制散点分布图供辅助说明.

已知  $U(10) = \{(x_1, y_1), \dots, (x_{167}, y_{167})\}$  中共有 167 个定子温度曲线散点. 首先, 按照定子温度由小到大升序排列, 使其满足  $y_i > y_{i-1}, i \in (2, 167)$  的必要条件; 然

后, 依次计算前  $i$  个坐标点中定子温度的方差值, 通过对文献的了解, 将计算结果形象地称作滑差值, 所得到的数据集称为滑差集; 最后, 通过与初始阈值  $S$  比较分析, 从而实现数据清洗的功能. 为保持数据完整性和参考性, 这里定义第 1 个坐标点对应的滑差值为 0; 另外, 为了缩小数据值范围, 以下计算均用标准方差值代替方差值, 处理后的 10 号区间数据如表 1 所示.

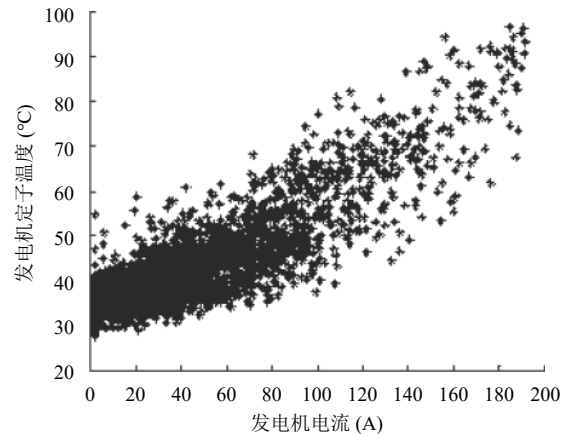


图 1 定子温度曲线图

表 1 区间 10 处理后的部分数据

新序号	发电机电流 (A)	定子温度 (°C)	滑差值	原序号
1	51.615 67	33.939	0	1742
2	51.8955	34.403	0.026 912	28
3	50.813 67	34.775	0.081 048	1743
4	51.660 83	35.572	0.283 492	1745
⋮	⋮	⋮	⋮	⋮
166	54.3275	58.992	34.6729	3801
167	53.0075	59.541	36.108 06	1948

根据式 (3) 计算小于阈值  $S$  的最大  $\gamma$ , 通过定义的公式 (4) 得到 10 号区间的阈值  $S$  为 13.8504, 即计算小于 13.8504 的最大  $\gamma=84$ , 对应  $S$  为 13.885 86, 从而由此得到以 (84, 13.885 86) 为临界点, 可划分两个数据集. 如图 2 所示, 由 5、10、20 区间分别通过这种方法共同绘制的散点图, 可以清晰地看到两种数据集的划分.

通过分析大量的测试数据后, 如图 3 所示, 是结合全部区间进行 OIV 算法的散点图, 可以通过灰度值的不同直观由  $S$  阈值划分的两种数据集.

### 1.4 对最优组方差算法的改进

根据上述的式 (3)、(4), 发电机电流范围 0~183 A, 以每 5 A 进行区域分隔, 其中  $T \geq 180$  为  $U(36)$  区域, 通过定义  $S < \bar{\delta}$  再对正常与非正常进行划分.

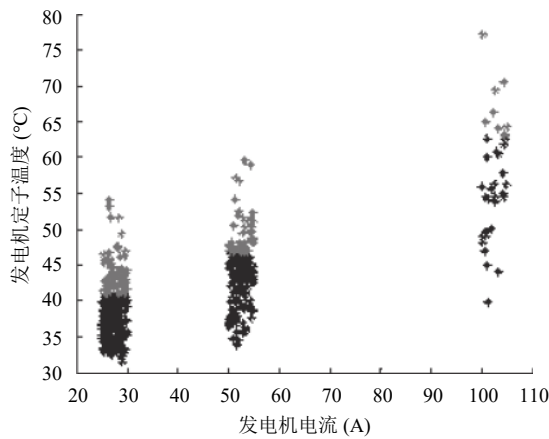


图2 定子温度5、10、20号分区图

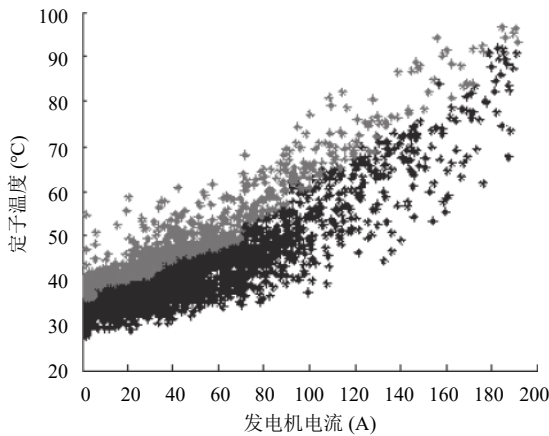


图3 时候定子温度曲线图

为了得到阈值  $S$  在这些数据中最合适的值, 结合图3发现图中散点的上限和下限都存在较为分散的数据点, 方差阈值  $S$  在原来 OIV 算法的基础上进行改进, 设  $n_1$ 、 $n_2$  为波动的百分比, 方差阈值  $S$  则定义为:

$$\bar{\delta} * (1 - n_1) < S < \bar{\delta} * (1 + n_2) \quad (9)$$

并对 OIV 算法式 (3) 改进为:

$$\gamma = \arg \max \left\{ S_1 < \sum_{\gamma=2}^n \left( \frac{1}{\gamma} \sum_{j=1}^{\gamma} (y_j - \bar{y}_{\gamma})^2 \right) < S_2 \right\} \quad (10)$$

通过以上推论得到主要改进的公式:

$$S_1 = \bar{\delta} * (1 - n_1) \quad (11)$$

$$S_2 = \bar{\delta} * (1 + n_2) \quad (12)$$

其中,  $i=1$  或  $2$ ,  $n_i \in (-1, 1)$ ,  $S_1$  与  $S_2$  即为整个数据集的方差阈值上下限.  $y_j$  为  $j$  个点的定子温度;  $\bar{y}_{\gamma}$  为第 1 到第  $\gamma$  个点的定子温度平均数,  $\gamma$  为常数.

## 2 清洗、建模与测试

### 2.1 环境温度作为第二输入参数

根据 SCADA 系统所采集的数据, 发电机电流是最能影响发电机定子温度的参数, 在未进行数据清洗的情况下, 输入参数只有发电机电流所测试的  $MSE$  结果为 21.6341 左右, 这时考虑到发电机外界因素环境温度的影响, 进行一次输入参数为发电机电流和环境温度所测试的  $MSE$  结果为 12.5506 左右, 相比单输入参数, 数据集在精度上得到了明显的提升.

### 2.2 与发电机电流和环境温度

根据上文中提到的方法和方差阈值  $S$ , 首先通过式 (10)、式 (11) 和式 (12) 调整整个方差阈值上限  $S_2$ , 同时暂时将  $S_1$  定为 0, 再根据  $0 < S < \bar{\delta} * (1 + n_2)$ , 为了确定最佳上下限幅值, 通过表 2 实验的结果可以看出, 上下限幅值不宜过高, 按照原定的 0~1(100%) 为最优上下限幅值.

表 2 设定上下限幅值后方差阈值实验结果

阈值上下限长度	一次 MSE	二次 MSE	三次 MSE	四次 MSE	五次 MSE
0~1	3.8314	3.6761	3.7108	3.6935	3.6280
0~1.1	3.9081	4.5272	4.3060	3.9397	4.6981
0~1.2	4.2503	4.1191	4.3205	4.0357	4.3008
0~1.3	4.8518	4.3167	4.3994	4.4342	4.3864
0~1.4	5.0346	4.8051	4.8108	4.9043	4.4789
0~1.5	5.1928	4.9257	5.1588	4.9357	4.7536

为了更进一步地降低  $MSE$  值, 开始表 3 的实验.

表 3 调整方差阈值  $S$  下限后的实验结果

$S$ 上下限长度	一次 MSE	二次 MSE	三次 MSE	四次 MSE	五次 MSE
0.1~1.1	2.6577	2.6662	2.6480	2.6873	2.7946
0.2~1.2	2.3946	2.4233	2.4260	2.3573	2.4715
0.3~1.3	2.4581	2.4056	2.4169	2.3637	2.3658
0.4~1.4	2.0363	2.1457	2.0684	2.0138	2.0071
0.5~1.5	2.1064	2.0630	1.9351	2.0140	2.0776

由表 3 可见当确定了方差阈值的下限, 整体的  $MSE$  得到了明显的下降, 该数据集的精度和使用可靠性也提高了不少.

### 2.3 测试改进 OIV 清洗后的效果

接下来需要对本文中的改进 OIV 算法进行测试, 首先从 SCADA 系统采集的数据进行算法数据清洗, 采用神经网络, 输入参数为清洗之后的相关特征量和前一时段的发电机定子温度, 输出为清洗之后的发电机温度参数.

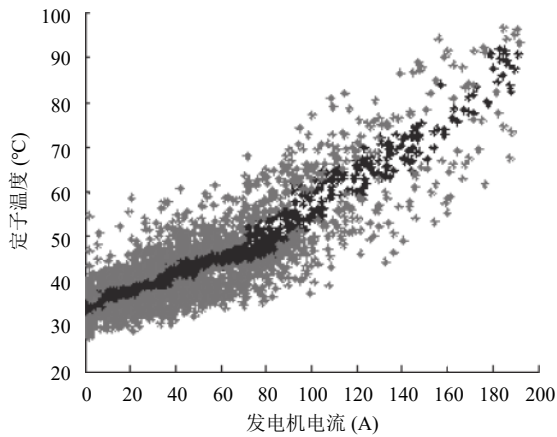


图4  $\bar{\delta} * 0.5 < S < \bar{\delta} * 1.5$ 所得到的发电机定子温度曲线图

根据式 (9)、(10)、(11)、(12) 和表 3 所提供的信息, 为了建立神经网络模型, 设  $n_1=n_2=0.5$ , 经过改进 OIV 算法, 得到图 4 所示, 提取正常数据集.

为了验证本文所提到的数据清洗方法 OIV 算法的精确度, 从风电场 SCADA 系统中下载了 2016 年 1 月 1 日 00 时 00 分 00 秒至 6 月 31 日 23 时 59 分 59 秒的数据进行实验验证, 其中包含了发电机电流、环境温度、发电机定子温度. 由于数据量过于庞大, 在使用前先按照 IEC61400-12 的标准进行十分钟平均化, 所得结果如表 4 所示, 并结合定子温度曲线来评估发电机定子的状态.

表 4 风电场 SCADA 数据经 10 min 平均化后的数据

发电机电流 (A)	环境温度 (°C)	定子温度
75.600 33	18.7698	38.417
95.593 16	18.9578	40.607
68.575 42	19.0865	41.115
...	...	...

将 2015 年 5 月的数据作为测试量使用原 OIV 算法和使用改进 OIV 算法的两种情况分别进行的模型测试实验结果见图 5, 见图 6.

对比图 5 和图 6 可以看出, 改进 OIV 清洗后得到的数据集所建立的模型在预测过程中, 可见改进 OIV 算法的清洗功能的精度高且可靠.

为了确定该模型的可行性, 将此模型进行实时预测, 实时使用的数据量为 2017 年 4 月到 6 月, 如图 7 可以看出, 预测值和真实值整体的趋势非常相似, 再通过如图 8 的误差值可以看出, 这实时三个月期间, 存在那么多处误差高于 2°C 的时刻, 并通过 SCADA

系统实时报备的信息对应, 温度偏高、定子异常、温度数据读取错误等异常点信息基本上——对应, 可见该方法可行.

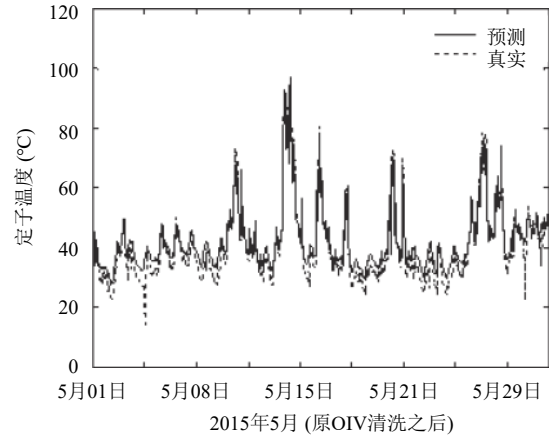


图 5 进行原 OIV 清洗直接做的测试结果

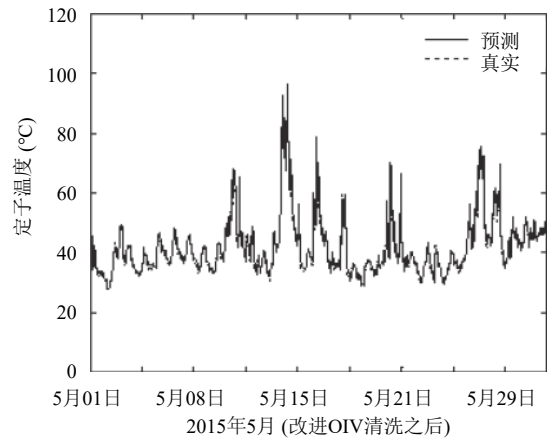


图 6 进行改进 OIV 清洗之后做的测试结果

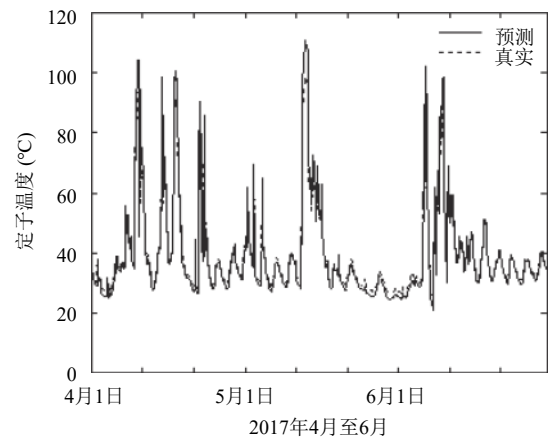


图 7 实时 2017 年 4 月至 6 月进行的预测波形图

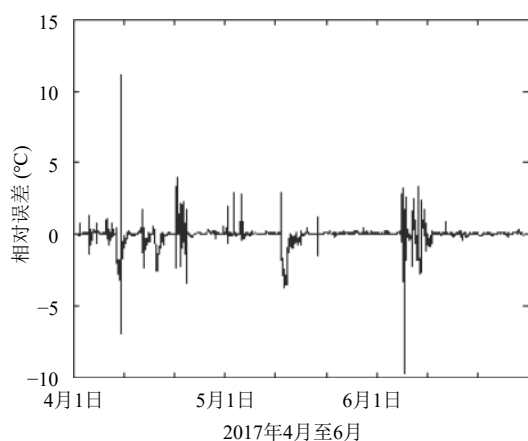


图8 实时预测误差结果图

### 3 结论

为解决发电机定子温度的预测前期数据预处理复杂的问题,利用 SCADA 系统采集相关数据,结合发电机理论中有关定子的温度参数,提出了一种基于风电场 SCADA 数据定子温度的预处理算法,并通过实验和改进得到了一个较为理想的方法,得到以下结论。

原有的最优组方差 (OIV) 算法可以清洗非正常数据集,提取正常数据集,但通过实验表明原有的方差阈值单向区域的定义和实验结果发现该算法在精度上还不够,故将此进行了双向改进,定义了  $S_1$  和  $S_2$ ,且通过实验证明了改进 OIV 更适用于定子温度预测数据预处理。对于现有的风电场的条件,结合 SCADA 系统采集的数据,本文所提到的改进最优组内方差算法 (OIV),计算过程简单易懂,通用性强。很好的解决了对于庞大的 SCADA 数据去差取优困难的问题,加快了整个数据预处理的过程,结合神经网络工具箱的测试功能,为之后的预测及故障诊断工作作出一个很好的基础。

#### 参考文献

- 1 李俊峰,蔡韦波,乔黎明,等. 2013 中国风电发展报告. 北京: 中国储能网, 2013.
- 2 陈雪峰,李继猛,程航,等. 风力发电机状态监测和故障诊断技术的研究与进展. 机械工程学报, 2011, 47(9): 45-52.
- 3 Lu B, Li YY, Wu X, *et al.* A review of recent advances in wind turbine condition monitoring and fault diagnosis. Proceedings of 2009 IEEE Power Electronics and Machines in Wind Applications. Lincoln, NE, USA. 2009. 1-7.
- 4 秦常贵. SCADA 系统及其在风力发电场的应用. 电力设备, 2004, 5(12): 31-33.
- 5 Feng YH, Qiu YN, Crabtree CJ, *et al.* Monitoring wind turbine gearboxes. Wind Energy, 2013, 16(5): 728-740.
- 6 Gray CS, Watson SJ. Physics of failure approach to wind turbine condition based maintenance. Wind Energy, 2010, 13(5): 395-405.
- 7 Popa LM, Jensen BB, Ritchie E, *et al.* Condition monitoring of wind generators. Proceedings of the 38th IAS Annual Meeting on Conference Record of the Industry Applications Conference. Salt Lake City, UT, USA. 2003, 3: 1839-1846.
- 8 Nandi S, Toliyat HA. Condition monitoring and fault diagnosis of electrical machines—a review. Proceedings of the 34th IAS Annual Meeting. Conference Record of the 1999 IEEE Industry Applications Conference. Phoenix, AZ, USA. 1999, 1: 197-204.
- 9 赵海森,刘晓芳,杨亚秋,等. 基于时步有限元分析的超高效电机定子槽形优化设计. 中国电机工程学报, 2011, 31(33): 115-122.
- 10 娄建楼,胥佳,陆恒,等. 基于功率曲线的风电机组数据清洗算法. 电力系统自动化, 2016, 40(10): 116-121.
- 11 马幼捷,杨海珊,周雪松,等. 基于神经网络的风电场建模. 中国电力, 2010, 43(9): 79-82.
- 12 代悦. 神经网络在火电厂建模中的应用[硕士学位论文]. 保定: 华北电力大学(保定), 2003.
- 13 Tallam RM, Habetler TG, Harley RG, *et al.* Neural network based on-line stator winding turn fault detection for induction motors. Proceedings of Conference Record of the 2000 IEEE Industry Applications Conference. Rome, Italy. 2000, 1: 375-380.