

基于改进 bin 算法的风电机组风速-功率数据清洗

王新, 王政霞

(重庆交通大学, 重庆 400074)

摘要: 风速-功率是风电机组发电性能的重要指标, 对风电场的运行管理具有重要意义。风速-功率数据是通过安装在风电场的监视控制与数据采集 (SCADA) 系统采集得到的, 原始数据存在大量噪声, 给后续应用研究带来了很大的挑战。基于风速-功率数据的空间分布特征, 将风速-功率数据分为 3 类, 并改进了数据预处理方法 bin 算法, 提出了基于分区域 (dbin) 算法的异常数据识别清洗方法及流程。实验结果表明, dbin 算法识别异常数据的效率比传统算法更高, 具有较强的通用性。

关键词: 风电机组; 风速功率; dbin; 监视控制与数据采集

中图分类号: TM315

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202007

Wind speed-power data cleaning of wind turbine based on improved bin algorithm

WANG Xin, WANG Zhengxia

Chongqing Jiaotong University, Chongqing 400074, China

Abstract: Wind power is an important indicator of the generating performance of wind turbines, which is of great significance to the operation and management of wind farms. The wind-speed and power data were collected through the monitoring and control (SCADA) system installed in the wind farm. There are a lot of noises in the original data, which brings great challenges to the subsequent application research. Based on the spatial distribution characteristics of wind-speed and power data, wind-speed and power data was divided into three categories, the data preprocessing method bin algorithm was improved, and the method and process of abnormal data identification and cleaning based on district bin (dbin) algorithm were proposed. The experimental results show that the dbin algorithm proposed in this paper is more efficient than the traditional algorithm in identifying abnormal data, and has strong universality.

Key words: wind turbines, wind-speed and power, dbin, supervisory control and data acquisition

1 引言

随着社会经济的不断发展, 传统能源的短缺问题已成为制约当今社会经济发展的首要问题。风能作为可再生的清洁能源, 储量巨大、分布广泛, 且开发利用成本较低, 是可再生能源中最具开发前景的一种能源^[1]。风能作为一种成熟的技术, 在过去的几十年中得到了快速的发展。但是在风力发电过

程中, 风具有高度随机性和波动性, 风速-功率也会变化, 这对电力系统有一定的影响。通过风力发电机组输出功率和风速得到的风速-功率曲线可以对风电机组的发电性能做出有效的评估, 同时, 风速-功率曲线也是预测风电场风速-功率的重要指标^[2]。通常, 风电机组运行的风速-功率数据是通过安装在风电场的监视控制与数据采集 (supervisory control and data acquisition, SCADA) 系统采集得

收稿日期: 2020-02-05; 修回日期: 2020-02-25

通信作者: 王政霞, zxiawang@163.com

基金项目: 重庆市科委自然科学基金资助项目 (No.2018jcyjAX0398); 重庆市研究生科研创新项目 (No.2019S0104)

Foundation Items: Natural Science Foundation Project of CQCSTC (No.2018jcyjAX0398), Graduate Student Research Innovation Project in Chongqing (No.2019S0104)

到的,但是,SCADA 系统采集到的运行数据中存在大量的异常数据。产生异常数据的原因主要有 3 种。一是风机多位于风能资源丰富的区域,其运行环境较为恶劣^[3],例如戈壁、荒野、高山等都会对风机造成一定的影响,导致传感器易产生故障,风电机组因检修故障计划停机也可能造成测量数据偏离正常值。二是由于 SCADA 系统的控制终端与风电场距离较远,数据传输过程中会产生噪声及外界电磁干扰,使运行数据产生异常。三是弃风限电^[4]现象,弃风限电是当地电网接纳能力不足和风电不稳定等导致的部分风机暂停的现象。弃风限电会导致风电机组的输出功率长期维持在一个较低的水平,即使风速超过额定风速,输出功率也会低于额定功率,这就会造成原始数据中存在大量的异常数据,这些异常数据会直接影响预测模型的建模精度。因此,为了提高运行数据的数据质量,有必要对 SCADA 系统采集到的运行数据进行数据清洗^[5],以提高运行数据的质量。

风电机组风速-功率数据清洗是风电机组异常运行数据识别与清洗中的研究热点,其中,将风电机组风速-功率数据清洗分为 3 类。第一类是基于数理统计的清洗方法(如密度、距离、中位数、方差、平均值等),但该类方法对分布密集的异常数据的识别效果有限,例如文献[6]采用四分位法对异常数据进行处理,但是在风电数据中异常数据所占比重较大,四分位法会去除大量正常数据;文献[7]采用四分位法与聚类分析法剔除异常数据,但没有给出具体的剔除标准,可能会导致正常数据被误删,对数据处理结果影响较大;文献[8]利用局部离群因子(local outlier factor, LOF)算法来检测异常值,该算法根据风速-功率曲线,计算曲线附近的相对密度,去除异常点。第二类是基于数据挖掘的方法,即在大量的数据中通过无监督的学习方法辨识特征表现异常的数据,用于数据分析处理,该类方法不需要事先建立一个模型,具有较好的抗干扰性,例如文献[9]采用 LOF 和支持向量机相结合的方法,通过 LOF 值固定异常范围,然后利用支持向量机建立判别机制,从而判定异常值,有效地避免了基于距离聚类方法的异常值淹没现象。文献[10]用最优方差算法实现了分散型异常数据的有效识别,但不能有效识别密度较高的堆积型异常数据。第三类方法基于数据的分布特征,该类方法的依据是位于风速-功率曲线之外的数据为异常数据,可实现多类型异常

数据的清洗,例如在假定风速-功率的概率密度误差服从正态分布的基础上,文献[11]提出了基于“ 3σ 法则”的判定方法来检测异常值,即位于 $[0, 3\sigma]$ 之外的误差为异常误差,位于 $[0, 3\sigma]$ 之内的误差为正常误差,从而达到数据清洗的目的。但是风速-功率的概率密度往往为多峰,精度不高。文献[12]利用成熟预测算法获取残差的误差信息,对误差概率密度特性进行了综合分析,并确定异常信息。文献[13]提出了一种不要求取概率密度分布函数的四分位算法,避免了拟合概率密度的复杂过程。

综上所述,基于数理统计和数据挖掘的方法对分布密集的异常数据的识别效果有限,且需要大量正常数据作为样本,普适性较差。本文提出的 dbin 方法根据异常数据的位置,将位于风速-功率曲线之外的数据作为异常数据,可实现多类型异常数据的清洗,而且不需要数据样本来训练,通用性强。

2 风速功率数据的分布特征

2.1 标准功率曲线建立

风速-功率曲线的建立需要风速-功率的历史数据,而风电机组标准的风速-功率曲线是由风电机组厂商依据 GB/T 18451.2-2012《风力发电机组 功率特性测试》绘制的。其中规定,标准空气密度为 1.225 kg/m^3 ,评估方法按照 10 min 内的统计平均值进行评估,即计算风电机组正常运行的每个 10 min 时间段内的平均风速、平均功率。某额定功率为 2.1 mW、切入风速为 3 m/s、额定风速为 11 m/s、切出风速为 22 m/s 的风电机组的标准风速-功率曲线如图 1 所示。

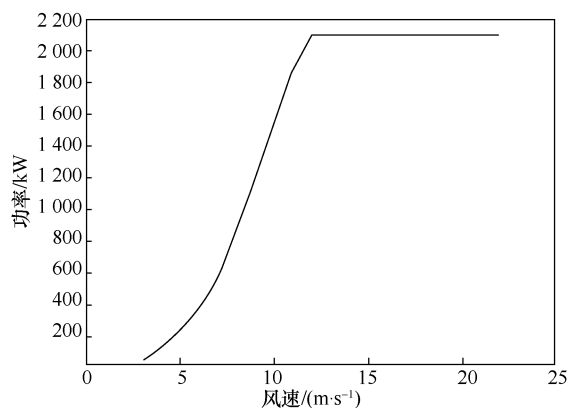


图 1 2.1 mW 风电机组的标准风速-功率曲线

在风速达到切入风速时,机组开始产生并输出功率;随着风速增加,输出功率也逐渐增加;风速达到额定风速时,机组输出功率达到额定值 2.1 mW;

当风速超过额定风速时,通过变桨控制可使输出功率稳定在限定值(即额定功率 2.1 mW)。

2.2 异常数据分布特征

从风电场采集到的数据中通常包含大量异常数据,这些异常数据会直接影响预测模型的建模精度,对风速-功率预测有一定的影响。按照数据点在标准风速-功率曲线上的分布特征,异常数据可分为3类:第一类为曲线底部连续时间内风速很大但功率为零的数据;第二类为曲线中部风速很大、功率很小或小于额定功率的数据,即弃风数据或故障数据;第三类为风速很小但功率很大的数据。这里选取国内某风场风电机组的运行数据进行验证,该风场风机的基本参数为:切入风速 3 m/s,切出风速 22 m/s,额定功率 2 100 kW,额定风速 11 m/s。这里选取异常数据分布比较典型的1号、2号、3号风机连续12个月的数据。1号风机的各类异常数据的分布如图2所示。

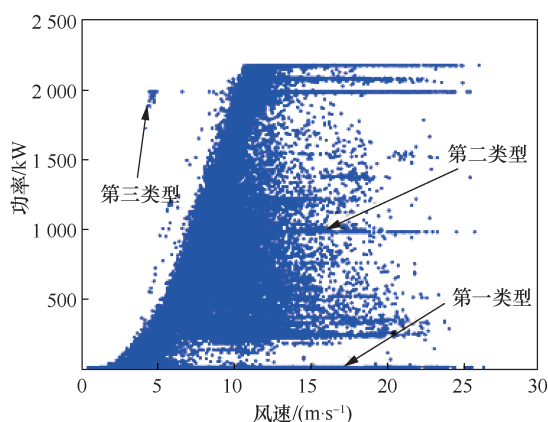


图2 1号风机风速-功率异常数据特征分布

第一类异常数据在风速-功率曲线中表现为风速很大、功率为零。此类异常数据产生的原因是传感器故障或风电机组因检修故障计划停机。在这种情况下,风力发电机的输出为零,但风机的检测系统仍需要工作,故而会产生负功率的数据^[14]。

第二类异常数据在风速-功率曲线中表现为一条或多条位于风速-功率曲线下界的横向密集数据带和密度较低的无规律散点。弃风限电和信号噪声、环境因素、极端天气是此类数据产生的主要原因,在风电场的实际运行过程中,由于目前电力系统的调峰调频能力和输电能力不足,强制弃风已成为常态^[15],这就使得风电机组的输出功率低于正常输出功率,在图2中表现为一条或多条横向密集数据带。

第三类异常数据为风速很小但功率很大的数

据,在风速-功率曲线中表现为多条功率曲线上界之外的横向密集数据带,产生的原因是风速传感器失灵或通信错误^[16]。

3 基于 district bin 算法的风速-功率曲线拟合

3.1 bin 算法

bin 算法^[17]用于对监测数据进行平滑处理,其基本原理是:在切入风速($V_{\min}$)与切出风速($V_{\max}$)之间划出若干个风速区间,将所有数据点按照其风速大小归入各个风速区间,然后对各区间内的数据点进行统计分析,得出各个监测参数的统计值,最终以这些风速区间内监测参数的统计值代替原始数据进行后续的数据分析工作。1号风机数据拟合结果如图3所示。

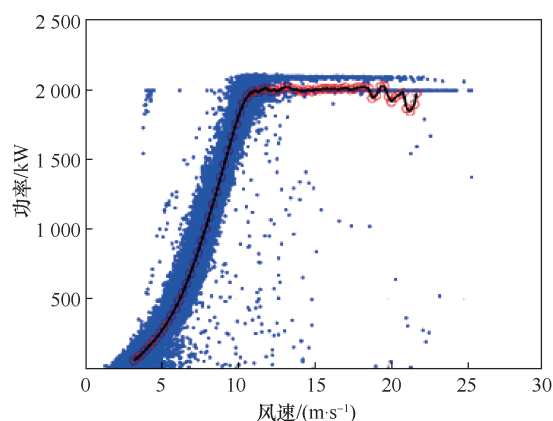


图3 1号风机 bin 方法拟合风速-功率曲线

bin 算法如下。

算法1 bin 算法

输入: 风速 $V=(v_1, v_2, \dots, v_n)$

功率 $P=(p_1, p_2, \dots, p_n)$

(1) 切入风速 $V_{\min} \in V$, 切出风速 $V_{\max} \in V$, 定义风速区间 $[V_{\min}, V_{\max}]$

(2) 划分为 N 个风速区间

$$N \approx \frac{V_{\max} - V_{\min}}{0.25}, N \text{ 为正整数}$$

(3) 令 $K=1$ 且 $K < N$

(4) 寻找位于区间 $[V_{\min}(K-1)V_{\text{step}}, V_{\min}(K+1)V_{\text{step}}]$ 的风速数据 VI , 其中 V_{step} 为步长,

$$V_{\text{step}} = \frac{V_{\max} - V_{\min}}{N}$$

(5) 对第 K 个子范围内的风速数据 VI 进行统计分析,即进一步将 $[\min(VI), \max(VI)]$ 划分为 m 个格,然后计算第 K 个子范围内的风速期望值 VI (见

式 (1))

(6) 同理计算出风速对应的功率

$$P_i = \sum_{j=1}^m (p'_j \cdot p_j) \quad (j=1,2,\dots,m)$$

(7) end

3.2 district bin 算法

bin 算法使用期望值描述单个 SCADA 数据, 减少了 SCADA 数据中由“异常值”引起的统计误差, 实现了数据规约、数据平滑和异常值剔除。然而, bin 算法没有考虑到在风速位于额定风速与切出风速之间时功率保持不变的情况。使用 bin 算法会导致风功率为额定功率时出现拟合不准的现象 (如图 3 所示)。本文在 bin 的基础上进行改进, 提出了使用分区域 (district bin) 算法 (即 dbin 算法) 进行划区域计算, 这样能有效避免风速-功率曲线在额定功率时出现拟合不准的现象 (如图 4 所示)。dbin 算法的基本原理是: 在切入风速 ($V_{\min}$)、额定风速 (V_r)、切出风速 ($V_{\max}$) 之间划出 $N1=[V_{\min}, V_r]$ 和 $N2=[V_r, V_{\max}]$ 两个风速区间, 将所有数据点按照其风速大小归入这两个风速区间, 然后仍使用 bin 算法对 $N1$ 风速区间内的数据进行统计分析, 基于式 (1) 和式 (2) 对 $N2$ 风速区间内的数据进行统计分析, 得出各个监测参数的统计值, 最终以这些风速区间内监测参数的统计值代替原始数据进行后续的数据分析工作。dbin 算法保留了 bin 算法实现的数据规约、数据平滑和异常值剔除的优点, 同时能够有效地避免风功率为额定功率时出现拟合不准的现象。1 号风机数据拟合结果如图 4 所示。

$$VI = \sum_{j=1}^m (p_j \cdot v_j) \quad (j=1,2,\dots,m) \quad (1)$$

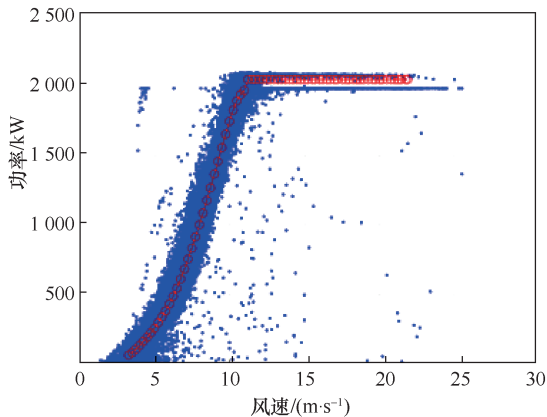


图 4 1 号风机 dbin 方法拟合风速-功率曲线

其中, $p_j = \frac{n_j}{n} \times 100\%$ 表示风速位于第 j 个 bin 中的概率; n_j 是位于第 j 个 bin 中的风速数据数量, n 是位于第 K 个子范围内的风速数据总数; v_j 是与第 j 个箱子相对应的风速。

$$P_i = p_r \quad (2)$$

其中, P_r 为额定功率。

dbin 算法如下。

算法 2 dbin 算法

输入: 风速 $V=(v_1, v_2, \dots, v_n)$

功率 $P=(p_1, p_2, \dots, p_n)$

(1) 切入风速 $V_{\min} \in V$, 切出风速 $V_{\max} \in V$,

额定风速 $V_r \in V$, 额定功率 $P_r \in P$

定义区间 $[V_{\min}, V_r]$ 和 $[V_r, V_{\max}]$

(2) 划分 $N1$ 、 $N2$ 个两个风速区间

$$N1 \approx \frac{V_r - V_{\min}}{0.25} \quad N2 \approx \frac{V_{\max} - V_r}{0.25}$$

$N1$ 、 $N2$ 为正整数

(3) for $K=1$ 且 $K < N$

(4) 分别寻找位于区间 $[V_{\min}(K-1)V_{\text{step1}}, V_{\min}(K+1)V_{\text{step1}}]$ 的风速数据 V_m 和 $[V_{\min}(K-1)V_{\text{step2}}, V_{\min}(K+1)V_{\text{step2}}]$ 的风速数据 V_n , 其中 V_{step1} 、 V_{step2} 为步长

$$V_{\text{step1}} = \frac{V_r - V_{\min}}{N1} \quad V_{\text{step2}} = \frac{V_{\max} - V_r}{N2}$$

(5) 对 $N1$ 区间第 K 个子范围内的风速数据 V_m 进行统计分析, 即进一步将 $[\min(V_m), \max(V_m)]$ 划分为 m 个格, 然后用式 (1) 计算第 K 个子范围内的风速期望值 VI

(6) 同理计算出风速对应的功率

$$P_i = \sum_{j=1}^m (p'_j \cdot p_j) \quad (j=1,2,\dots,m)$$

(7) 对 $N2$ 区间第 K 个子范围内的风速数据 V_n 进行统计分析, 即进一步将 $[\min(V_n), \max(V_n)]$ 划分为 m 个格, 然后用公式计算第 K 个子范围内的风速期望值 VI

(8) $N2$ 区间风速对应的功率为

$$P_i = p_r \quad (j=1,2,\dots,m)$$

(9) end

3.3 数据清洗流程

第 3.2 节的结果表明, dbin 算法可对风电机组的风速-功率曲线进行有效的拟合, 拟合结果更契

合厂商给出的标准风速-功率曲线, 这为后面的数据清洗提供了理论依据。通过分析风电机组的实际风速-功率曲线发现, 风速-功率运行数据中第二类堆积型异常数据居多。因此, 在进行数据清洗时, 首先要基于限电功率对原始异常数据进行清洗, 去除大部分第二类堆积型异常数据, 继而采用 dbin 算法对数据进行训练, 拟合参考功率曲线。将功率曲线拉直, 横坐标仍为实际运行的风速, 纵坐标为实际的运行功率与在参考功率曲线中对应的参考功率的差值, 实际运行功率越接近参考功率, 这个差值越小。根据风机运行数据统计差值分布的标准差 σ , 位于 3σ 内的数据为正常数据, 否则为异常数据, 从而达到清洗原始数据的目的。具体的数据清洗流程如图 5 所示。

4 实验结果与分析

4.1 数据样本

选取国内某风场风电机组的运行数据进行验证, 该风场风电机组的基本参数为: 切入风速 3 m/s, 切出风速 22 m/s, 额定功率 2 100 kW, 额定风速 11 m/s。这里选取异常数据分布比较典型的 1 号、2 号、3 号风机连续 12 个月的数据来进行验证。由图 6 (a) ~ 图 6 (c) 可知, 1 号、2 号、3 号风机连续 12 个月的原始数据中存在大量的第二类异常数据。基于历史

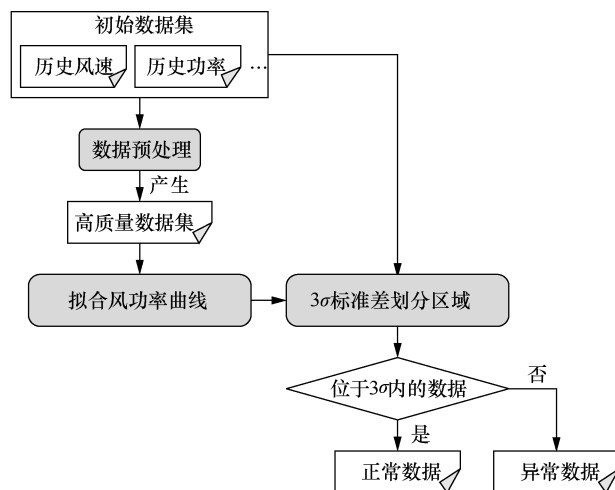


图5 数据清洗流程

数据中的限电数据对原始数据进行初步预处理, 得到对应的高质量数据集, 但仍存在大量的异常数据, 如图 6 (d) ~ 图 6 (f) 所示。

4.2 风速-功率曲线拟合

根据风机设计的设计标准风速-功率曲线, 风机的运行数据分布在设计曲线两侧。继而分别采用 dbin 算法、bin 算法、支持向量回归 (support vector regression, SVR) 算法、反向传播 (back propagation, BP) 算法对数据进行训练, 并拟合参考功率曲线, 由图 7 ~ 图 9 可以看出, dbin 算法的拟合曲线最贴近标准风速-功率曲线。

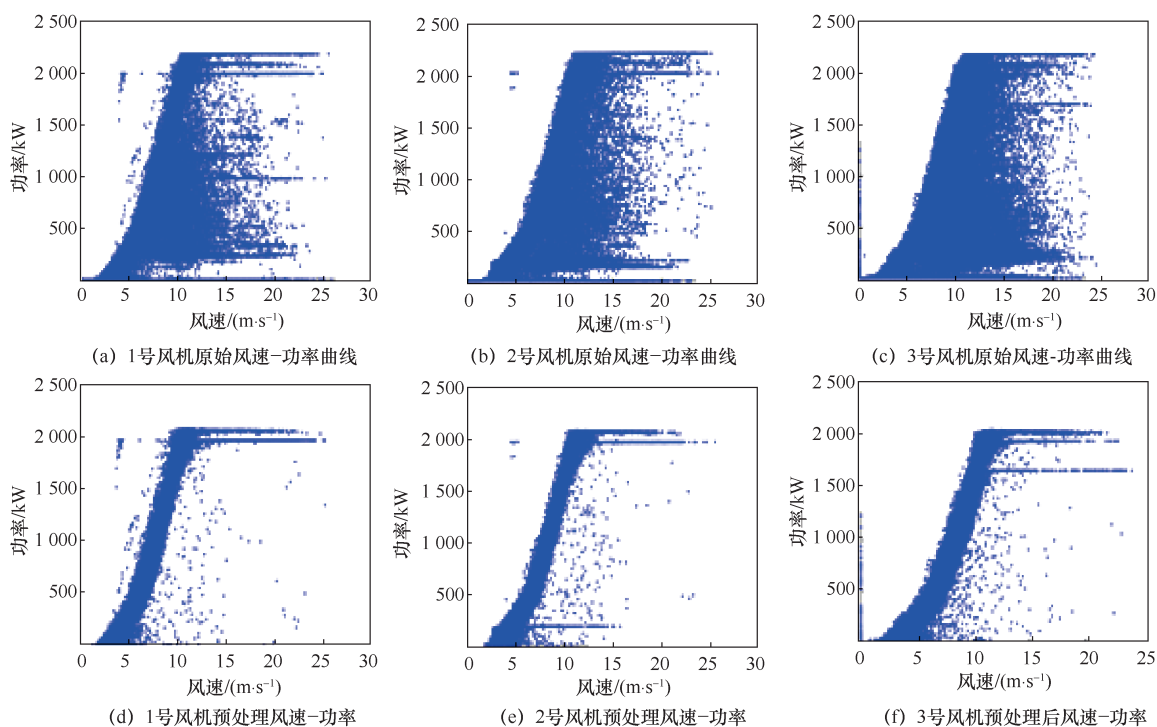


图6 风机风速-功率数据集

(1) 1号风机风速-功率曲线拟合

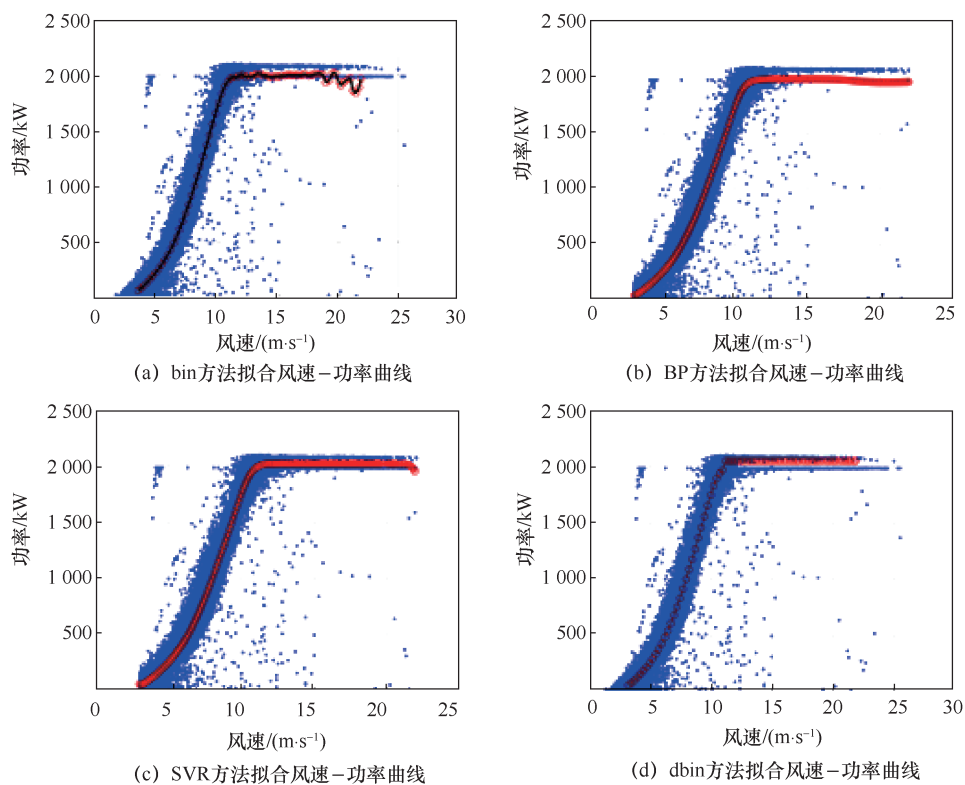


图7 1号风机风速-功率曲线拟合

(2) 2号风机风速-功率曲线拟合

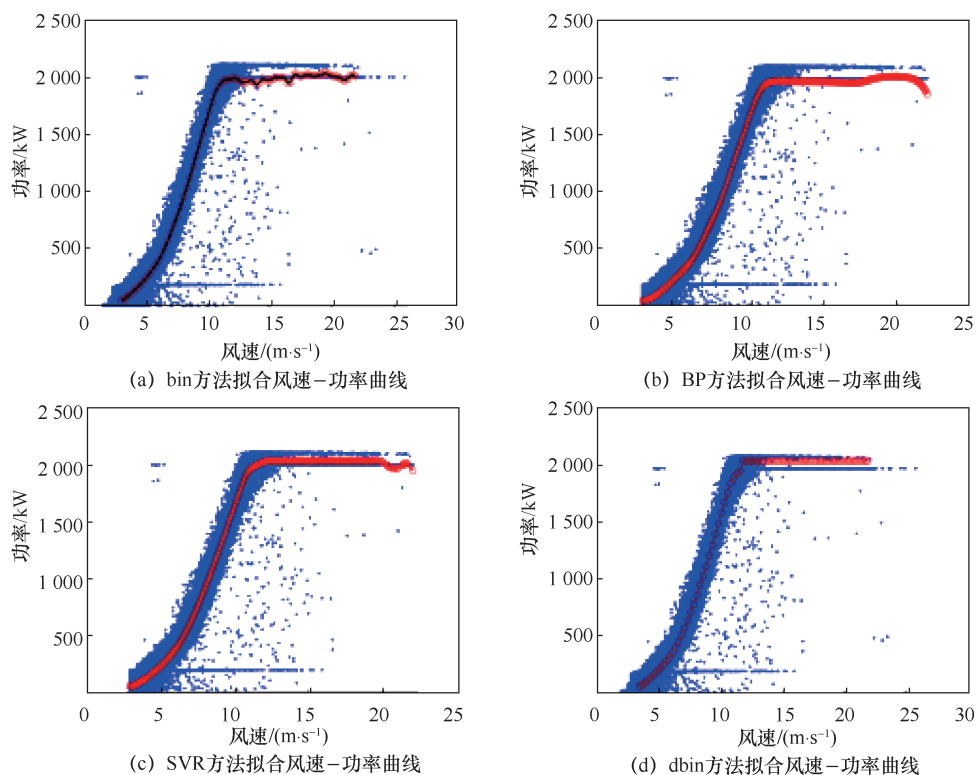


图8 2号风机风速-功率曲线拟合

(3) 3号风机风速-功率曲线拟合

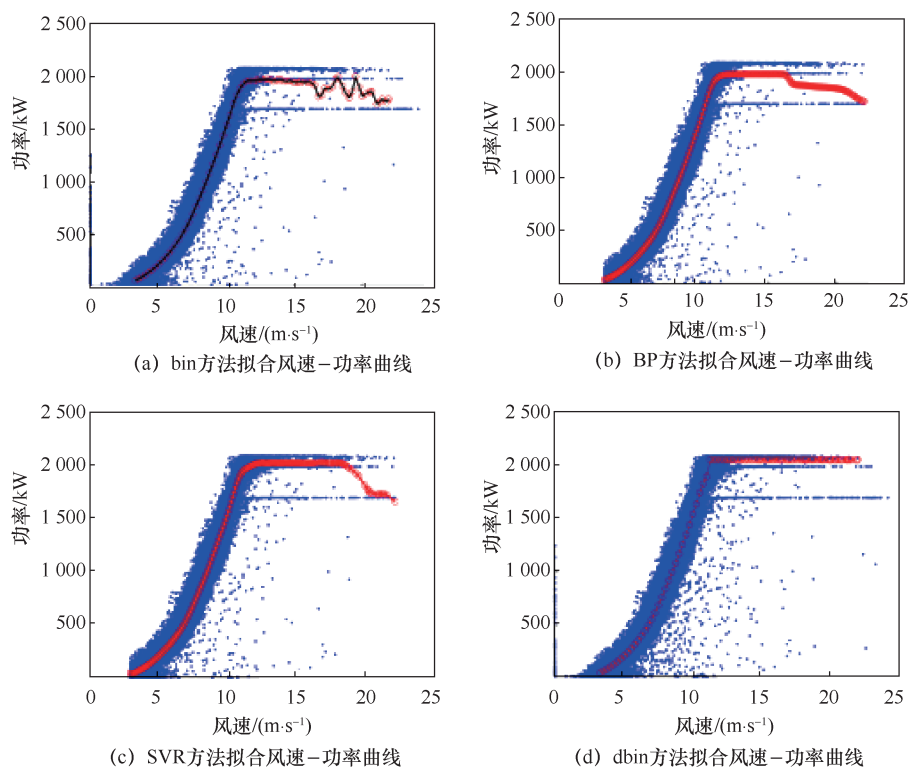


图9 3号风机风速-功率曲线拟合

4.3 3 σ 划分

将功率曲线拉直, 横坐标仍为实际运行的风速, 纵坐标为实际的运行功率与在参考功率曲线中对应的参考功率的差值, 实际运行功率越接近参考功率曲线, 这个差值越小。根据风机运行数据统计差值分布的标准差 σ , 位于 3σ 内的数据为正常数据, 其余为异常数据。由图10可知, bin算法中大部分数据位于 3σ 之外, 不能进行有效地数据清洗; BP算法中, 预测的数据明显小于实际风速, 不符合预测标准; SVR算法能够较好地进行预测且数据位于 3σ 之内, 但其得到的预测值误差范围波动较大, 不易于进行数据清洗; dbin算法中大部分数据位于 3σ 之内, 能够有效地对数据进行清洗。

4.4 效果验证

图11~图13分别为dbin算法、bin算法、SVR算法、BP算法对3个机组异常数据的清洗结果, 其中蓝色为各风机原始数据, 红色为处理后的正常数据。如图11~图13所示, dbin算法能够对提出的3类异常数据进行有效清洗, 清洗后的数据符合标准风速-功率曲线的分布特点; bin算法

也能够对提出的3类异常数据进行有效的清洗, 但bin算法也会对正常数据进行清洗, 不符合标准风速-功率曲线分布特点; BP算法和SVR算法对第二类异常数据的识别效果不佳, 不能达到数据清洗的目的。

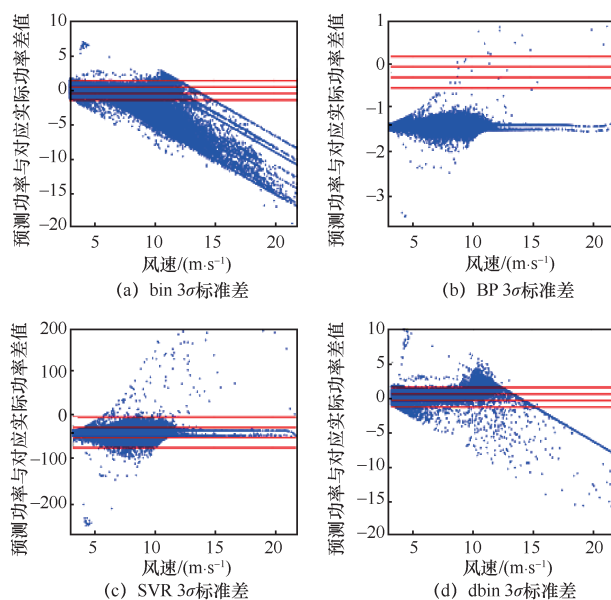


图10 1号风机参考功率与实际功率差值分布

(1) 1号风机数据清洗

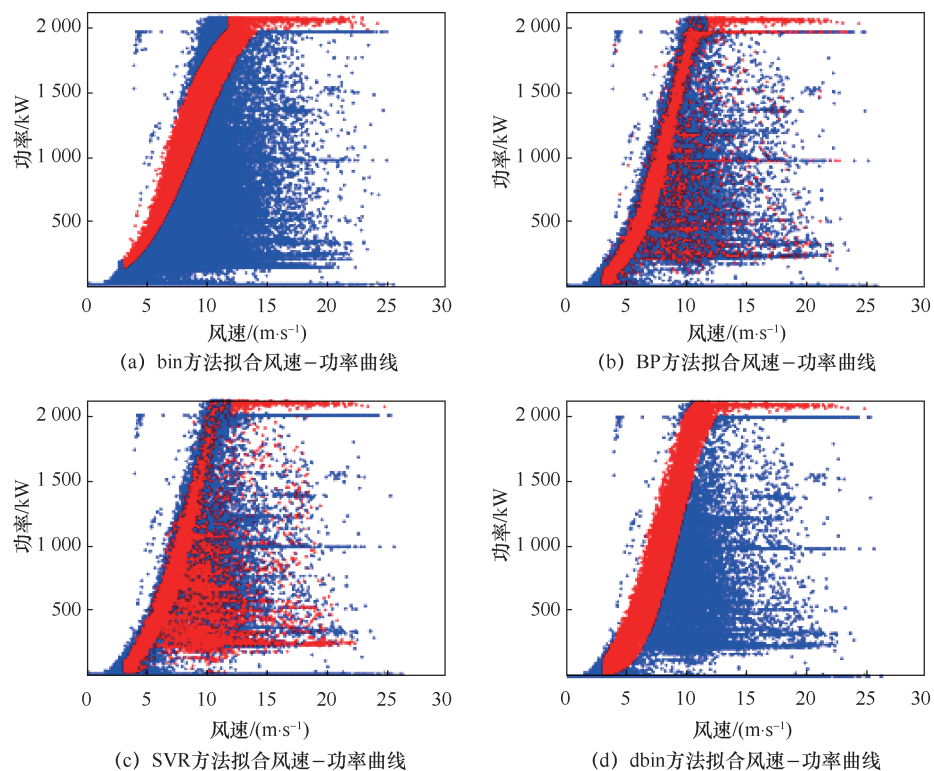


图 11 1号风机数据清洗结果

(2) 2号风机数据清洗

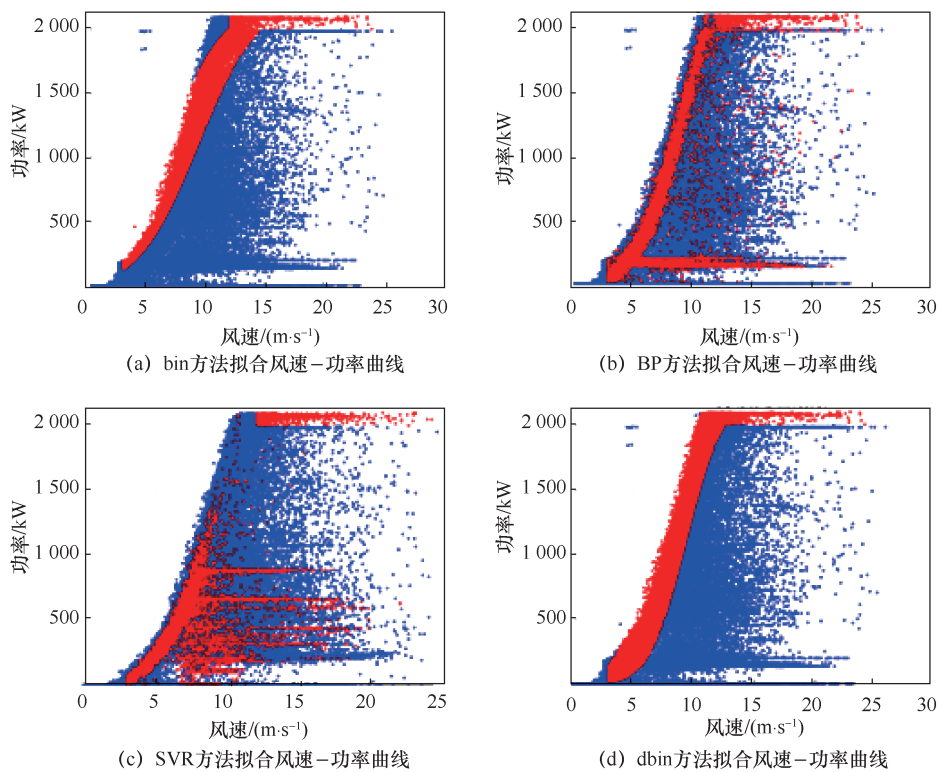


图 12 2号风机数据清洗结果

(3) 3号风机数据清洗

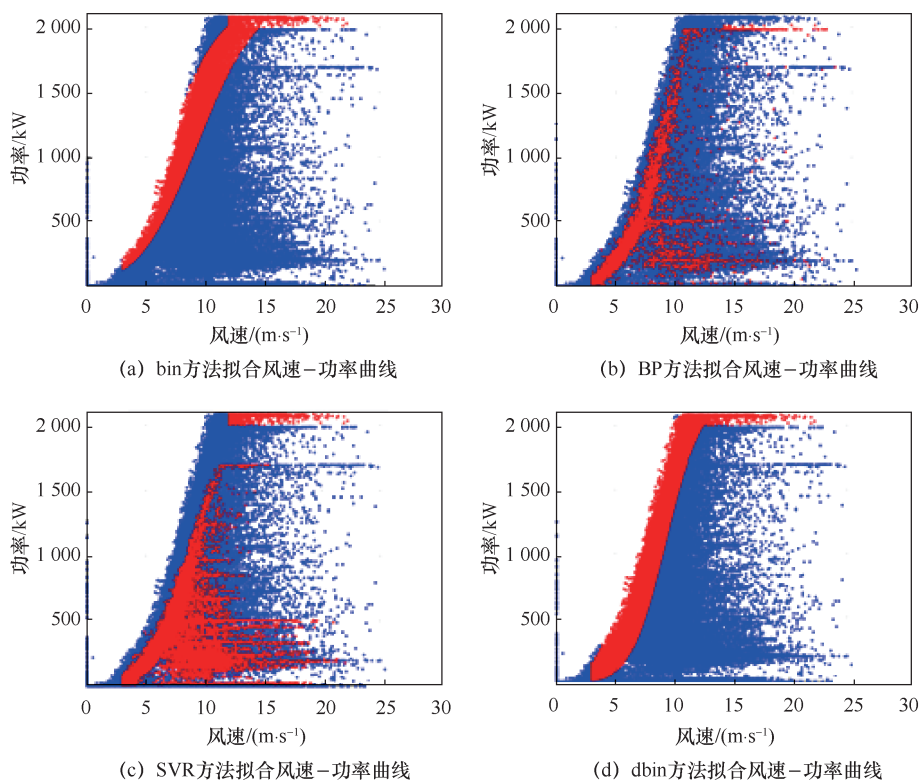


图13 3号风机数据清洗结果

4.5 算法对比分析

表1记录了dbin算法、bin算法、SVR算法、BP算法对3个机组的数据删除率和清洗效率。bin算法能够去除异常数据，但风功率为额定功率时出现拟合不准的现象，会导致正常数据被删除，且其数据删除率达到了75.4%；BP算法和SVR算法对第二类异常数据的识别效果不佳，数据删除率分别为66.8%和56.7%，且SVR算法的处理时间较长；dbin算法对数据的清洗效果最好，能对所有类型的异常数据进行有效识别，且清洗后的数据更加符合标准风速-功率曲线的分布特征。此算法对各类异常数据的删除率均在40.9%左右，耗时约为3.4s（这与异常数据的多少有关），清洗效率较高，具有较高的可行性。

综上所述，dbin算法为风电机组风速-功率运行数据的清洗提供了高效、合理、可靠的方法。dbin算法能实现对风速-功率曲线中堆积型异常数据和分散型异常数据的有效识别，可满足不同机组、不同异常数据分布状况的清洗需求，特别是当机组的风速-功率数据中存在大量堆积型异常数据时，dbin算法具有明显的优势。

表1 不同算法的数据清洗效果

清洗方案	机组编号	原始数据/条	剩余数据/条	数据删除量	清洗时间/s
bin	1	156 919	49 272	68.6%	3.38
	2	175 806	40 895	76.7%	3.40
	3	230 817	41 313	82.1%	3.73
BP	1	156 919	43 745	72.1%	6.33
	2	175 806	60 515	65.6%	6.250
	3	230 817	85 930	62.8%	10.03
SVR	1	156 919	84 853	45.9%	854.64
	2	175 806	62 457	64.5%	171.37
	3	230 817	93 024	59.7%	670.33
dbin	1	156 919	102 415	34.7%	3.38
	2	175 806	98 765	43.8%	3.53
	3	230 817	128 463	44.3%	3.43

5 结束语

风速-功率数据中通常包含大量异常数据，这些异常数据会直接影响预测模型的建模精度，因而对电力系统有一定的影响。本文依据标准风速-功率曲线，提出了基于dbin算法的数据清洗方案，并

与传统的 bin 算法、BP 算法、SVR 算法进行了对比。实例表明, 基于 dbin 算法的清洗流程能够有效识别异常数据, 效率较高。

参考文献:

- [1] 胡锦涛, 朱峙成, 刘佳玮, 等. 基于大数据的风电场风能资源及其利用情况评估[J]. 科技创新导报, 2018, 15(16): 48-49.
HU J N, ZHU Z C, LIU J W, et al. Evaluation of wind energy resources and utilization in wind farms based on large data[J]. Science and Technology Innovation Herald, 2018, 15(16): 48-49.
- [2] 杜保华, 王曦, 范奇, 等. 基于风能利用率的风电场运维水平评价方法[J]. 热力发电, 2019, 48(7): 117-121.
DU B H, WANG X, FAN Q, et al. Evaluation method for operation and maintenance level of wind farm based on wind energy availability rate[J]. Thermal Power Generation, 2019, 48(7): 117-121.
- [3] 张运洲, 胡泊. “三北”地区风电开发、输送及消纳研究[J]. 中国电力, 2012, 45(9): 1-6.
ZHANG Y Z, HU B. Study on wind power exploitation, transmission and accommodation in northern region of China[J]. Electric Power, 2012, 45(9): 1-6.
- [4] 郭鹏, 文晶, 朱丹丹, 等. 基于源-荷互动的大规模风电消纳协调控制策略[J]. 电工技术学报, 2017, 32(3): 1-9.
GUO P, WEN J, ZHU D D, et al. The coordination control strategy for large-scale wind power consumption based on source-load interactive[J]. Transactions of China Electrotechnical Society, 2017, 32(3): 1-9.
- [5] 吴亚联, 梁坤鑫, 苏永新. 基于 SCADA 数据驱动的风电机组部件故障预警[J]. 无线互联科技, 2018, 137(13): 128-133.
WU Y L, LIANG K X, SU Y X, et al. Fault warning of wind turbine components based on SCADA data[J]. Wireless Internet Technology, 2018, 137(13): 128-133.
- [6] 朱倩雯, 叶林, 赵永宁, 等. 风电场输出功率异常数据识别与重构方法研究[J]. 电力系统保护与控制, 2015, 43(3): 38-45.
ZHU Q W, YE L, ZHAO Y N, et al. Methods for elimination and reconstruction of abnormal power data in wind farms[J]. Power System Protection and Control, 2015, 43(3): 38-45.
- [7] 赵永宁, 叶林, 朱倩雯. 风电场弃风异常数据簇的特征及处理方法[J]. 电力系统自动化, 2014, 38(27): 39-46.
ZHAO Y N, YE L, ZHU Q W. Characteristics and processing method of abnormal data clusters caused by wind curtailments in wind farms[J]. Automation of Electric Power Systems, 2014, 38(27): 39-46.
- [8] ZHENG L, HU W, MIN Y. Raw wind data preprocessing: a data-mining approach[J]. IEEE Transactions on Sustainable Energy, 2015, 6(1): 11-19.
- [9] 胡伟, 李勇, 曹一家, 等. 基于 LOF 和 SVM 的智能配电网故障辨识方法[J]. 电力自动化设备, 2016, 36(6): 7-12.
HU W, LI Y, CAO Y J, et al. Fault identification based on LOF and SVM for smart distribution network[J]. Electric Power Automation Equipment, 2016, 36(6): 7-12.
- [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[J]. 电力系统自动化, 2016, 40(10): 116-121.
LOU J L, XU J, LU H, et al. Wind turbine data-cleaning algorithm based on power curve[J]. Automation of Electric Power Systems, 2016, 40(10): 116-121.
- [11] 李九龙, 周凌柯. 基于“ 3σ 法则”的显著误差检测[J]. 计算机与现代化, 2012(1): 10-13.
LI J L, ZHOU L K. Detecting and identifying gross errors based on “ 3σ rule”[J]. Computer and Modernization, 2012(1): 10-13.
- [12] 叶林, 任成, 赵永宁, 等. 超短期风电功率预测误差数值特性分层分析方法[J]. 中国电机工程学报, 2016, 36(3): 692-700.
YE L, REN C, ZHAO Y N, et al. Stratification analysis approach of numerical characteristics for ultra-short-term wind power forecasting error[J]. Proceedings of the CSEE, 2016, 36(3): 692-700.
- [13] 毛李帆, 姚建刚. 中长期负荷预测的异常数据辨识与缺失数据处理[J]. 电网技术, 2010, 34(7): 148-153.
MAO L F, YAO J G. Abnormal data identification and missing data filling in medium-and long-term load forecasting[J]. Power System Technology, 2010, 34(7): 148-153.
- [14] 范晓泉, 杜大军, 费敏锐. 风电异常测量数据智能识别方法研究[J]. 仪表技术, 2017(1): 10-14.
FAN X Q, DU D J, FEI M R. Research on the intelligent identification method for abnormal measurement data of the wind power[J]. Instrumentation Technology, 2017(1): 10-14.
- [15] 田书欣, 程浩忠, 曾平良, 等. 基于调频层面的风电弃风分析[J]. 电工技术学报, 2015, 30(7): 18-26.
TIAN S X, CHENG H Z, ZENG P L, et al. Analysis on wind power curtailment at frequency adjustment level[J]. Transactions of China Electrotechnical Society, 2015, 30(7): 18-26.
- [16] YE X, LU Z, QIAO Y, et al. Correction of outliers in wind farm identification and time series data[J]. IEEE Transactions on Power Systems, 2016, 31(6): 4197-4205.
- [17] YANG W X, COURT R, JIANG J K. Wind turbine condition monitoring by the approach of SCADA data analysis[J]. Renewable Energy, 2013, 53(9): 365-376.

[作者简介]



王新 (1991-), 男, 重庆交通大学硕士生, 主要研究方向为机器学习与数据挖掘、故障检测。



王政霞 (1977-), 女, 重庆交通大学教授、硕士生导师, 主要研究方向为机器学习与数据挖掘、图像处理、医学影像诊断。