

基于云计算平台的风电场 SCADA 系统的设计与应用^①

梁 涛^② 侯振国^③ 邹继行 张迎娟

(河北工业大学控制科学与工程学院 天津 300130)

摘 要 针对目前风电场积累的海量多源异构的大数据,借助云计算平台架构灵活多变、数据并行处理的优势,设计了基于云计算平台的风电场数据采集与监控(SCADA)系统。在该系统的云计算平台上采用智能化数据挖掘算法,完成了风机数据的远程采集、智能分析、故障预判。最后选取反映风电场性能的功率和转子转速数据,以自适应权重(AFW)的 K-means 算法和最小二乘支持向量机两种典型的数据挖掘算法为例,对数据进行了聚类分析和回归分析,验证了系统的可行性。

关键词 数据采集与监控(SCADA),云计算,数据挖掘,聚类,支持向量机(SVM)

0 引 言

伴随着我国风电技术的逐步成熟,风电产业迅猛发展,我国已拥有全球规模最大的风电市场^[1]。数据采集与监控(supervisory control and data acquisition, SCADA)系统的出现改变了风电场的运维模式,系统以“无人或少人值守、远程监控”为目标^[2],设立远程集控中心,工作人员摆脱了恶劣的工作环境,降低了风电场运维成本。

但是风电场 SCADA 系统所采集的大量的数据,未被充分利用和挖掘,而只是停留在实时数据监控、历史数据显示、报表统计等方面^[3]。同时异构的风电大数据造成历史库指数级的增长,造成数据中心面临瘫痪风险,导致 SCADA 系统开始面临计算机 CPU 升级、内存不足、成本增加等一系列的问题,文献[1]实现了风电场 SCADA 系统完整的功能,但是系统采用传统的数据处理方式,难以应对风电大数据,传统的数据处理方式的弊端显现出来。

伴随着机器学习与数据挖掘技术尤其是大数据与智能化技术的兴起,网络信息化时代到来,以互联网为基础的云计算平台的建设^[4],为风电企业的智能化发展提供了一种崭新的思路。文献[2]设计了基于云平台大数据的风电集控系统,实现了数据的实时采集和上传,但是没有做进一步的数据挖掘。本文设计了基于云计算平台的风电场远程集控中心系统,采用主流的 Hadoop + MapReduce 的分布式数据处理技术^[5,6],在该平台上充分利用风电机组的大量数据,运行智能化风机状态监控与故障诊断软件,做到故障预检预修,降低风机运维成本^[7]。通过实验验证,取得了较好的效果。

1 系统整体设计

风电场云计算平台 SCADA 系统如图 1 所示。系统将风电场 SCADA 系统与云计算平台相结合,发挥云计算平台的独特优势,使风电大数据的作用最大化。

① 河北省科技计划(16214510D)资助项目。
② 男,博士,教授;研究方向:自动控制、风力发电等;E-mail: 54008214@qq.com
③ 通讯作者,E-mail: 542984654@qq.com
(收稿日期:2017-07-07)

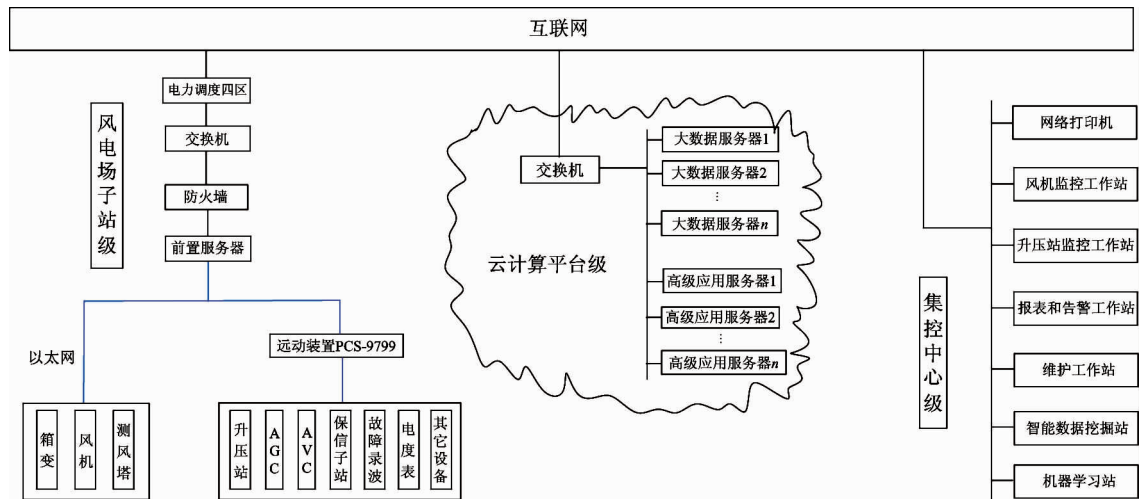


图1 风电场云计算平台 SCADA 系统图

该系统分为风电场子站级、云计算平台级和集控中心级三级。其中风电场子站级包括:风电场升压站设备如测风塔、电度表、自动发电控制(AGC)、自动电压控制(AVC)、保信子站、故障录波等设备,通过以太网和远动装置 PCS-9799 连接,再通过以太网和前置服务器连接;风机、箱变、测风塔等遥测、遥信、遥脉数据,通过以太网直接和前置服务器连接,数据传输协议采用 OPC 协议或者 modbus 协议。为了保证电力系统数据安全,前置服务器中的数据需要配置前置单向隔离装置以及防火墙,并采用“横向隔离,纵向加密”的原则组织网络架构,最后通过电力调度四区安全接入 Internet 互联网并上传到云计算平台级进行存储和处理,集控中心级通过用户访问接口接收云计算平台服务器处理后的运算结果和实时信息,集中展示在大屏幕上,供监控人员根据实际情况作出辅助决策和最终决断。同时,系统开发了手机移动客户端,方便工作人员随时随地查看自己所管辖风电场的运行情况,为企业开展区域规模化检修维护、合理优化资源配置、提高生产管理效率打下坚实的基础^[2]。

2 风电场云计算平台搭建

基于以上基础架构,搭建风电场云计算平台的三个主要部分:一是配套的云基础设施资源;二是风电场底层的数据采集^[8],风电场云计算平台具备数

据储存功能,储存风机、升压站设备、箱变设备、测风塔、电度表、AGC、AVC 和其他风电场需要的数据;三是智能数据挖掘算法和故障诊断专家系统策略,同时在数据库中还添加了调度管理、统计分析与报表、风场故障知识系统、物资管理等功能,为风电场运维人员工作提供强大的可视化数据支持。

云计算平台是一个大数据处理平台,其内部逻辑如图 2 所示,平台采用 Hadoop + MapReduce 的分布式数据处理技术^[5,6],与传统的数据处理方式相比,该平台基于灵活多变的开源框架,根据需要随时改变组件,且支持水平扩展,具有互联网属性、更加开放安全等明显优势。

3 系统功能

3.1 数据采集和传输

风电场侧数据采集的对象包括风机、升压站设备、测风塔以及箱变等,采集的变量包括遥测量、遥信量、遥控量、遥调量和电度量^[1]。升压站设备以及其他设备数据通过不同的通信规约经过以太网传送到远动装置 PCS-9799,经过规约转换,转换成标准的 104 规约数据包,经过网络防火墙,通过以太网传送到前置采集服务器;风机数据经过网络防火墙,通过以太网直接传送到前置采集服务器^[9]。数据采集和传输期间网络连接中断,会对数据采集产生很大影响,但此系统接口程序会重复检测网络连接

状态,断网时间段内,数据形成暂时的缓存文件,一旦网络连接恢复,立刻正常传输。

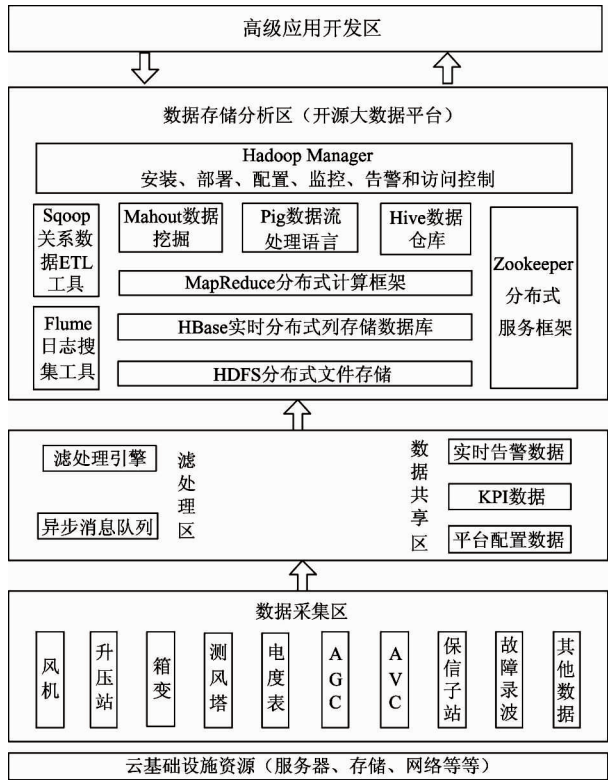


图 2 风电场云计算平台内部逻辑图

3.2 数据分布式云存储

风电场云计算平台具备数据存储功能,结合风电大数据和云计算平台的特点,采用图 3 所示的存储架构。

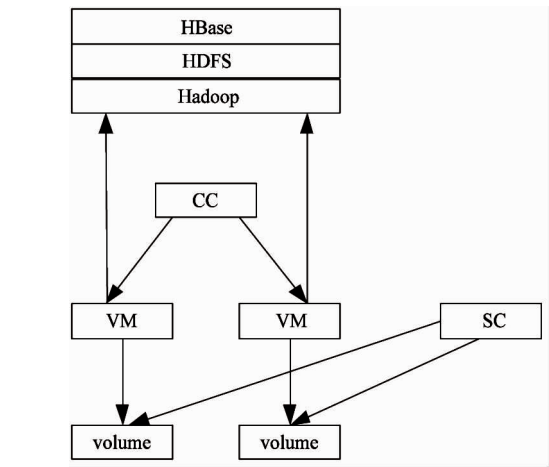


图 3 云计算平台存储架构图

此存储架构中存储控制器(SC)是部署在一台服务器上,与搭建的虚拟机(VM)服务器连接,VM

的数量可变,每一个 VM 还要关联一个 volume 来扩展存储,VM 之间还可以实现数据共享,共同访问一个大的数据存储区域^[10]。VM 与集群控制器(CC)连接,最终连接基于 Hadoop 基础框架的 HBase 数据库^[11]。

3.3 数据云计算

云计算平台的核心是数据处理,Hadoop 是与云计算平台最契合的分布式并行的运行架构,支持各种数据算法,包括数据排序、查询、图形分析、聚类分析、统计分析、最优化、数据挖掘、调度等等^[8]。风电场云计算平台是以风电大数据为输入,按照算法给定的规则,处理给定的数据,输出对应的结果。

Hadoop 架构的原理是将输入的海量数据划分成不同的区,原来的总任务被分割成几个小任务,每个小任务实例并行处理自己对应的分区数据。HBase 采用了基于列存储模式,能够便捷地为数据库中数据提供物理相邻的存储单元,因此能够快速读取和存储海量数据,采用 HBase 技术大大降低了搭建大规模结构化存储对硬件的要求。hive 是 Hadoop 平台的数据仓库工具,利用 hive 这个工具,能够把云计算平台中结构化的数据文件映射为数据库表,hive 工具还可以把 sql 语句转换为 MapReduce 任务分步执行。MapReduce 具有简单易理解、灵活多变、高容错的特点,是各种大数据处理算法都可以套用的并行处理的标准编程模型^[10]。根据风电大数据的特点,本文采用了主流的 Hadoop + MapReduce 的分布式数据处理方法。风电场云计算平台数据处理流程如图 4 所示。

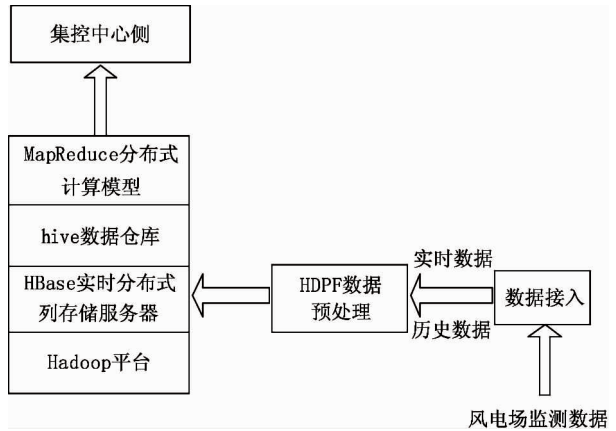


图 4 风电场云计算平台数据处理流程图

3.4 智能数据挖掘与故障识别算法的实现

风电场 SCADA 系统通常采用数值越限报警,报警模式单一,不具备预防故障恶化的功能,而大量的 SCADA 监测数据中隐含了风电机组各部件或子系统之间相互作用的关系。为了充分利用风机系统的 SCADA 海量数据,降低风电机组运行风险和减少风力发电的运行维护成本,需利用智能数据挖掘与机器学习算法来解决大型风电机组在线状态评估及故障诊断问题^[12]。基于云计算平台的数据挖掘架构如图 5 所示。因此,需在原有的 SCADA 报警系统上集成聚类、支持向量回归、神经网络、主元成分分析等算法。监控人员可通过 SCADA 系统的报警信息和聚类预测模型的预测值与实际监测值的偏差及趋势对风电机组的运行状态进行评估,还可以预判故障的发展趋势^[13]。本文选取了自适应权重的 K-means (AFW-K-means) 算法和最小二乘支持向量机 (least square support vector machine, LS-SVM) 两种典型的数据挖掘算法,来验证系统智能数据挖掘功能的实现。

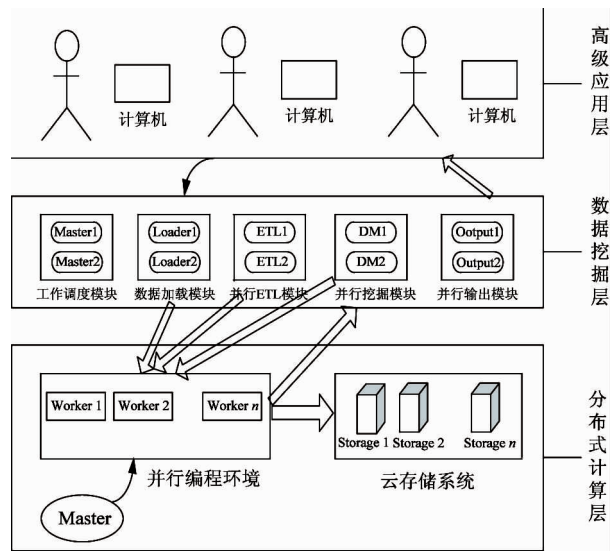


图 5 基于云计算平台的数据挖掘架构图

3.4.1 AFW-K-means 聚类算法的实现

K-means 聚类算法原理简单、理论成熟、收敛性强,是最经典的一种聚类算法,但是该算法最大的缺点就是初始聚类中心选取具有随机性。为了提高数据聚类的准确率,本文选取一种自适应权重的 K-

means 聚类算法即 AFW-K-means 算法,可以很好地弥补传统 K-means 算法的不足。

AFW-K-means 算法的原理是:首先计算对象在各属性上的值相对于均值的离散程度,然后以均方差和聚类个数 K 构造偏移因子,最后根据均值和偏移因子计算得到初始聚类中心。

下面对该算法作具体描述:

1. 输入输出

输入: $n \times m$ 数据集 (n 和 m 分别为对象个数和属性个数), 聚类个数 k ; 输出: k 个聚类,使每个对象到其所在聚类中心的加权欧式距离最小。

2. 具体步骤:

(1) 计算对象在各属性上的均值 $mean$ 和均方差 σ , 结果均为 $1 \times m$ 的行向量。

(2) 构造初始聚类中心 C :

$$C = \begin{cases} \left(mean \pm \frac{2\sigma}{k-1} \times j, j = 1, \dots, k/2 \right) \cup (mean), & k \text{ 为奇数} \\ \left(mean \pm \frac{2\sigma}{k} \times j, j = 1, \dots, k/2 \right) \cup (mean), & k \text{ 为偶数} \end{cases}$$

以上两式中的 $2\sigma/k$ 和 $2\sigma/(k-1)$ 称为偏移因子。

(3) 特征权重初始化: $\omega_j = 1/m, j = 1, 2, \dots, m$ 。

(4) 计算每个数据对象到聚类中心的加权欧式距离,按照距离最小原则为每个对象分配聚类号,重新计算聚类中心。

(5) 判断是否存在对象个数为 0 的聚类,如果是,说明数据对象分布非常密集,偏移因子过大。将偏移因子减半,重新选取聚类中心。

(6) 根据迭代结果,按照属性类内紧密、类间远离的原则调整每个属性的特征权重。

(7) 重复执行 (4) 和 (6) 直至达到预定的迭代数或每个聚类不再变化为止^[14]。

MapReduce 是云计算平台中一种标准的编程模型,其主要优势是,通过“Map (映射)”和“Reduce (归约)”步骤,使编程人员在不会分布式并行编程的情况下,将自己的程序运行在分布式系统上。当前的软件实现是指定一个 Map (映射) 函数,用来把

一组键值对映射成一组新的键值对,指定并发给 Reduce(归约)函数,用来保证所有映射的键值对中的每一个共享相同的键组^[15]。

在 AFW-K-means 聚类算法执行过程中,在计算 $D[0]$ 与 $cp[0] \cdots cp[k-1]$ 的距离的同时,可以计算 $D[1]$ 与 $cp[0] \cdots cp[k-1]$ 的距离,因此推断 AFW-K-means 算法与云计算平台结合 MapReduce 化是具有可行性的。

AFW-K-means 算法 MapReduce 化的执行步骤如下:

(1) 随机选择 k 个初始聚类中心,如 $cp[0] = D[0], cp[k-1] = D[k-1] \cdots$, 同时将这些初始聚类中心复制到 $OriginalCluster[]$ 中,并将 $OriginalCluster[]$ 分块,根据计算节点集群的情况,将 $OriginalCluster[]$ 块分配给各个计算节点;

(2) Map: 对于 $D[0] \cdots D[n]$, 分别计算其与 $cp[0] \cdots cp[n-1]$ 的距离,距离最近者记为 $c[i]$, $c[i_c]$ 的总个数记为 C_i , 同时在 MapReduce 框架下,把键值对 Key-Value 的 Key 和 Value 分别对应到 i 、 $D[k]$;

(3) Reduce: 由于 i 是 MapReduce 框架中键值对 Key,这保证了同一个 Key 的所有 $D[k]$ 会分配到同一个 Reduce 进程,则在此 Reduce 进程可以计算新的聚类中心 $cp[i] = (\sum c[i] \text{ 对应的 } D[j] / C_i)$, 并将此新的聚类中心存入 $DestinationCluster[]$;

(4) 比较 $DestinationCluster[]$ 与 $OriginalCluster[]$, 如果两者的变化小于预先给定的阈值,则聚类完成,否则将 $DestinationCluster[]$ 复制到 $OriginalCluster[]$ 后跳转到第 2 步继续执行。

3.4.2 最小二乘支持向量机算法的实现

最小二乘支持向量机 (LS-SVM) 是一种遵循结构风险最小化 (SRM) 原则的核函数学习机器,它是支持向量机 (SVM) 的一种特殊类型。最小二乘支持向量机是基于将支持向量机的不等式约束转化为求解线性方程组的算法,该算法不仅可以简化求解过程,而且可以提高求解收敛速度。其具体算法如下:

设给定的样本训练集为 (x_i, y_i) , 其中 $i = 1, 2, \cdots, l, x \in R^l, y \in R_{o \times x_i}$ 为第 i 个输入向量, y_i 为第

i 个输出向量,用非线性映射 $\varphi(\cdot)$ 将样本集输入空间 R^l 映射到特征空间 $\phi(x_i)$ 中,则在 $\varphi(x_i)$ 中构造最优决策函数:

$$y = w^T \varphi(x) + b \quad (1)$$

式中, w^T 为特征空间权系数向量, b 为偏置。根据结构风险最小化原则,可将最优问题表述为

$$\min \frac{1}{2} w^T w + c \sum_{i=1}^l e_i^2 \quad (2)$$

$$s. t. y = w^T \varphi(x_i) + b + e_i (i = 1 \sim l)$$

为求解上述问题,将约束优化问题转化为无约束优化问题,引入 Lagrange 函数,将上式中的优化问题变到对偶空间,得到下列式子:

$$L(w, b, e, \alpha) = \frac{1}{2} w^T w + c \sum_{i=1}^l e_i^2 - \sum_{i=1}^l \alpha_i [w^T \varphi(x_i) + b + e_i - y_i] \quad (3)$$

式中, α_i 为 Lagrange 乘子, e_i 为误差, c 为调节因子。由库恩-塔克条件可得

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \gamma e_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0 \end{cases} \quad (4)$$

消去参数 w 和 e_i , 则可以得到

$$\begin{bmatrix} 0 & e_l^T \\ e_l & Q + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (5)$$

其中, $e_l = [1, 2, \cdots, l], \alpha = [\alpha_1, \alpha_2, \cdots, \alpha_l]^T, y = [y_1, y_2, \cdots, y_l]^T, Q = \varphi(x_i)^T \varphi(x_i)$ 。由正定矩阵定理可以将 $Q = \varphi(x_i)^T \varphi(x_i)$ 的内积用核函数 $k(x_i, x_j)$ 代替,综上所述,可以得到 LS-SVM 的回归函数为

$$f(x) = \sum_{i=1}^l \alpha_i k(x_i, x_j) + b \quad (6)$$

LS-SVM 常用的核函数种类较多,在这些核函数中,向基核函数的结构简单,泛化能力好,其表达式为 $k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$, σ 为核宽

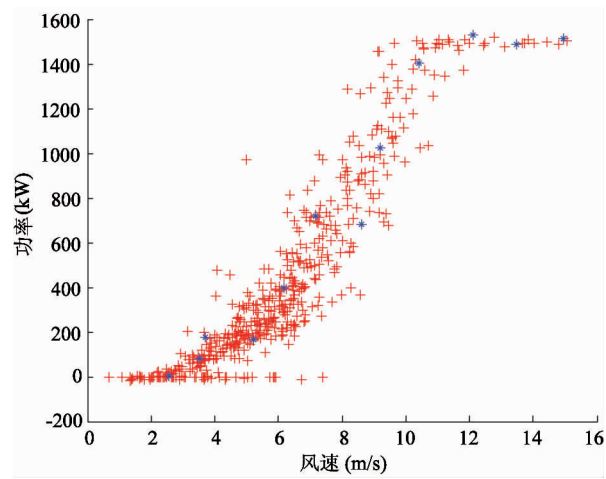
度。
基于此,本模型的核函数选择的是径向基函数。

4 仿真结果与分析

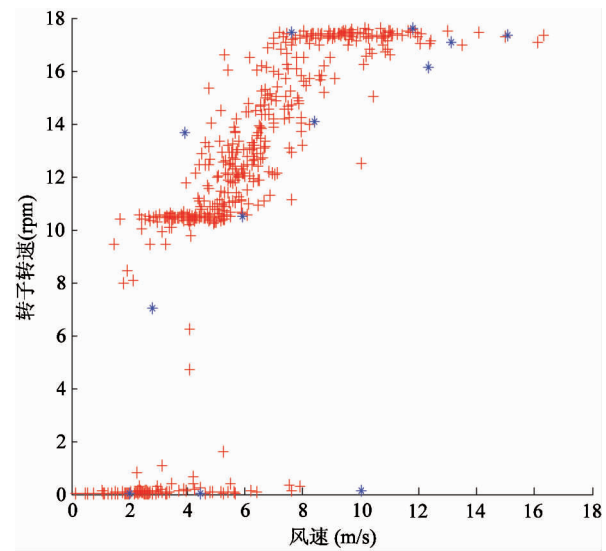
选取某风电场一定时间段内的功率和转子转速数据做数据挖掘实验,验证系统的可行性。

4.1 AFW-K-means 聚类算法

本文以选取三个聚类中心为例,通过网格把风速等分为 6 份,每一份选取三个聚类中心,每个聚类中心选取都有实际的意义,仿真结果如图 6 所示。



(a) 风电场功率散点图



(b) 风电场转子转速散点图

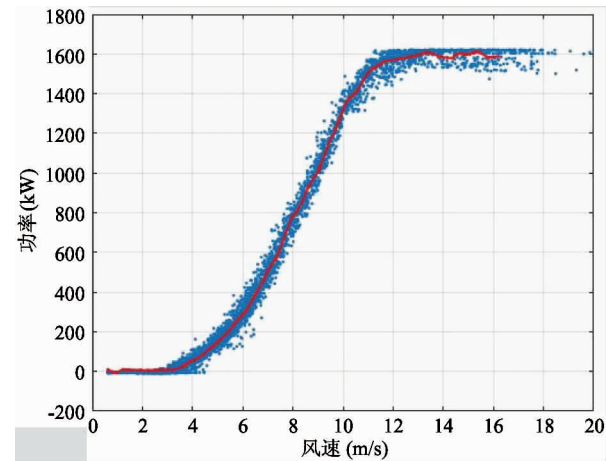
图 6 风电场功率和转子转速散点图

通过仿真可以看出,风机实际运行的功率和转

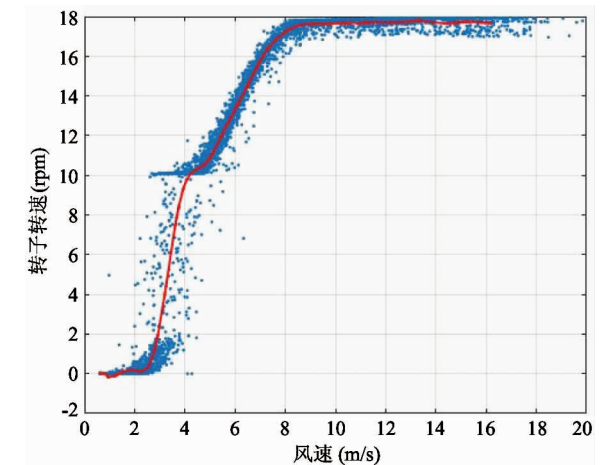
子转速的数据在三个聚类中心进行了聚类,假设我们给三个聚类中心分别编号 1、2、3,1 号聚类中心代表风机正常运行,2 号聚类中心代表一类故障,3 号聚类中心代表另一类故障,我们就可以对每个风速段内的风机运行情况进行评估。

4.2 最小二乘支持向量机算法

应用最小二乘支持向量机算法得到的功率曲线和转子转速曲线图如图 7 所示。



(a) 风电场功率曲线图



(b) 风电场转子转速曲线图

图 7 风电场功率和转子转速曲线图

通过风电场功率和转子转速的曲线图,我们可以对风机进行回归分析,根据预测模型的预测值与实际监测值的偏差及趋势对风电机组的运行状态进行评估,提高了状态评估的精度,还可以追踪和预判故障的发展趋势。

5 结 论

本文开发设计了基于云计算平台的风电场 SCADA 系统,利用云计算平台架构灵活多变、数据并行处理优势,在降低系统成本的同时,极大地提升了数据处理的速度。在云计算平台通过自适应权重的 K-means 算法和最小二乘支持向量机两种典型的数据挖掘算法,实现了风电场大数据的聚类分析和回归分析,验证了系统的可行性,为风电场将来的智能化发展提供了一种崭新的思路。

参考文献

- [1] 叶剑斌,左剑飞,黄小鉢. 风电场群远程集中 SCADA 系统设计[J]. 电力系统自动化,2010,34(23):97-101
- [2] 丛智慧. 基于云平台大数据技术的风电集控系统的设计与应用[J]. 内蒙古科技与经济,2016,(19):61-63
- [3] 赵国群,徐劲松. 基于 Web 的风电场监控及信息管理系统的研究与设计[J]. 机电工程,2014,31(12):1623-1628
- [4] 吴可心,李正民,张兆心,等. 云平台虚拟化性能的评测与研究[J]. 高技术通讯,2016,26(12):960-968
- [5] Sandholm T, Lai K. MapReduce optimization using regulated dynamic prioritization[J]. *Acm Sigmetrics Performance Evaluation Review*, 2009,37(1):99-310
- [6] Bu Y, Howe B, Balazinska M, Emst M D. The Hadoop approach to large-scale iterative data analysis[J]. *The*

VLDB Journal-The Int'l Journal on Very Large Data Bases, 2012,21(2):169-190

- [7] 卢锦玲,绳菲菲,赵洪山. 基于相关向量机的风机齿轮箱故障诊断方法[J]. 华北电力大学学报,2017,44(2):91-96
- [8] 罗贤缙,岳黎明,甄成刚. 风电场数据中心 Hadoop 云平台作业调度算法研究[J]. 计算机工程与应用,2015,51(15):266-270
- [9] 梁涛,袁正彬,梅春晓,等. 风电场群远程集中监控系统设计及智能化管理[J]. 自动化与仪表,2015,11:50-53
- [10] 钟运琴,方金云,赵晓芳. 大规模时空数据分布式存储方法研究[J]. 高技术通讯,2013,23(12):1219-1229
- [11] Zhao Y R, Wang W P, Meng D, et al. TDWS: A job scheduling algorithm based on mapReduce[C]. In: *Proceedings of the IEEE, International Conference on Networking, Architecture and Storage*, Xiamen, China, 2012. 313-319
- [12] 王意洁,孙伟东,周松,等. 云计算环境下的分布存储关键技术[J]. 软件学报,2012,23(4):962-986
- [13] 殷毅,任凯,曹阳. 基于改进的 MapReduce 模型的 Web 挖掘[J]. 科学技术与工程,2013,13(5):1206-1209
- [14] 李四海,满自斌. 自适应特征权重的 K-means 聚类算法[J]. 计算机技术与发展,2013,23(6):98-105
- [15] Yang H, Luan Z, Li W, Qian D. MapReduce workload modeling with statistical approach[J]. *Journal of Grid Computing*, 2012,10(2):279-310

Design and application of a wind farm SCADA system based on cloud computing platform

Liang Tao, Hou Zhenguo, Zou Jihang, Zhang Yingjuan

(School of Control Science and Engineering, Hebei University of Technology, Tianjin 300130)

Abstract

Aiming at the large amount of multi-source and heterogeneous data accumulated by wind farms, a wind farm supervisory control and data acquisition (SCADA) system based on the cloud computing platform is designed with the help of cloud computing platform's advantages of architecture flexibility and parallel data processing. The intelligent data mining algorithm is adopted in the cloud computing platform of the system to complete the remote collection, intelligent analysis and fault prediction of wind power generator's data. Finally, the power and rotor speed data reflecting the performance of the wind farm are selected and are analyzed by cluster analysis and regression analysis by using the two typical data mining algorithms of the adaptive weight based AFW-K-means and the least square support vector machine (SVM) to verify the feasibility of the system.

Key words: supervisory control and data acquisition(SCADA), cloud computing, data mining, clustering, support vector machine(SVM)