

## State Evaluation System of Mine Rotating Machinery Based on Big Data Mining

LI Xinhui\*

(Shaanxi energy Fengjiata mining operation Co., Ltd Yulin Shaanxi 719400 China)

**Abstract:** Intelligent state assessment and predictive maintenance of equipment is one of the core elements of building intelligent energy. Based on rotor dynamics, modern signal processing and artificial intelligence technology, a state evaluation system of mine rotating machinery based on historical big data mining and artificial intelligence algorithm is developed. In this system, the distributed data collector is used to build the condition monitoring network of the auxiliary machinery. Through the multi-source information deep fusion analysis technology, the intelligent evaluation of the state of the mine rotating machinery is realized.

**Key words:** state assessment; trend prediction; artificial intelligence; big data

EEACC: 7210      doi: 10.3969/j.issn.1005-9490.2021.02.032

## 基于大数据挖掘的矿井旋转机械状态评估系统

李新虎\*

(陕西能源冯家塔矿业运营有限责任公司 陕西 榆林 719400)

**摘 要:** 设备的智能化状态评估和预测维修是构建智慧能源的核心要素之一。以转子动力学、现代信号处理和人工智能技术为基础,开发了基于历史大数据挖掘和人工智能算法的矿井旋转机械状态评估系统。该系统采用分布式数据采集器构建辅机状态监测网,通过多源信息深度融合分析技术,实现了矿井旋转机械状态的智能评估。

**关键词:** 状态评估; 趋势预测; 人工智能; 大数据

中图分类号: TK38

文献标识码: A

文章编号: 1005-9490(2021)02-0434-05

当前,矿井旋转机械的检修维护多采用技术人员定期赴现场展开检测的方式进行,依靠人工在现场发现缺陷或检查故障,存在现场工作量大、故障发现概率低的缺点,并且因为缺乏对设备的实时监视预警能力,往往还会造成检修预案不全面,由小故障恶化为大故障的后果。

一方面,随着设备使用需求以及设备使用年限的增加,设备出故障的概率也越来越高,单纯依靠缩短检修周期将会给企业造成极大负担。另一方面,互联网技术迅速发展,电子设备越来越智能化,如何利用互联网技术与人工智能技术,开展设备自动化运检的相关研究,提高装置运行控制与保障能力,已经成为企业亟待解决的课题。

依托大数据技术运用关联规则算法,综合矿井机械设备的现场数据、历史运行数据(含故障分析)等信息,采用深度学习算法,建立设备的运行模型及故障预警知识库,从而对包括使用信息、调度信息、

计量点管理等设备在线运行状态给予全面评估,实现矿井旋转机械的动态运行分析与故障诊断。

### 1 基于大数据驱动的机械状态评估

#### 1.1 设备状态评估

设备运行状态的评估包括故障诊断和劣化趋势预测2部分。即不仅仅需要对当前的健康状态进行诊断,还需要确定部件或系统的剩余寿命和正常工作时间<sup>[4]</sup>。振动分析是矿井故障诊断的常用方法之一。振动参数在很大程度上表征了旋转机械的运行状态,但仅有振动而忽略机组的其他运行参数,将难以全面准确把握机组的运行状态和劣化趋势,容易造成故障的漏诊,无法满足设备智能化诊断和预测检修的需要。本文采用数据驱动的方法,通过对原始信号时域、频域和时频域特征提取,建立基于深度学习的故障诊断和劣化趋势预测模型,以实现设备状态的在线评估。

## 1.2 大数据重构与挖掘

基于大数据分析的故障诊断可以在收集到复杂装备运行特征数据的基础上,应用聚类、决策树等机器学习算法对大数据进行知识挖掘,获得与故障有关的诊断规则,从而实现对复杂装备的故障预测和诊断。

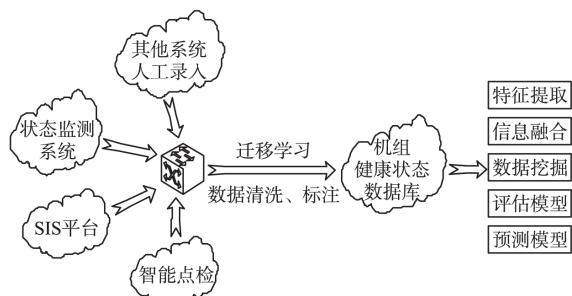


图1 大数据挖掘与重构

## 1.3 多源信息深度融合技术

不同类型的数据特征复杂度不同,数据流形结构不同,有效特征维度不同,对不同类型故障的敏感程度也不一样,简单的数据融合无法起到提高识别效率的目的,反而可能会因信息冗余和维数过高导致识别失效<sup>[6]</sup>。通过建立振动\温度\压力\开度\电量等状态参数深度融合的多视角特征学习模型,利用不同类型特征对故障的敏感差异来优化融合效果,可以有效解决不同视角特征的差异性和关联性互补融合问题,增强模型的鲁棒性能,提高设备状态评估及预测的准确性<sup>[7]</sup>。

## 2 状态评估算法

通过对机组历史运行数据,如风门开度、电机电流、轴承温度、振动、转速等的清洗和重构,在此基础上进行了大数据分析和机器学习建模。该模型经过训练和评价后,将在线运行,实时采集相应测点的数据,并利用训练后的模型进行在线分析、评价和预测。

### 2.1 状态聚类分析

聚类分析是一种广泛应用的无监督学习方法,在没有类别标记的情况下,可以提取数据中潜在的规则。聚类分析将样本分为几个不相交的子集,每个子集称为“簇”。 $K$ 均值聚类算法在给定聚类数 $k$ 之后,能够将数据集分成 $k$ 个“簇” $C = \{C_1, C_2, \dots, C_k\}$ ,算法优化的目标函数为样本的平方误差:

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - u_i\|^2 \quad (1)$$

算法的具体流程如下:

先随机选取 $K$ 个对象作为初始的聚类中心。然后计算每个对象与各个种子聚类中心之间的距离,把每个对象分配给距离它最近的聚类中心。聚类中心

以及分配给它们的对象就代表一个聚类。一旦全部对象都被分配了,每个聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是以下任何一个:

(1) 没有(或最小数目)对象被重新分配给不同的聚类。

(2) 没有(或最小数目)聚类中心再发生变化。

(3) 误差平方和局部最小。

图2给出了某一次设备运行状态 $K$ 均值聚类分析后的散点图,不同的颜色代表不同的状态,共5种(分别为正常状态,故障发展区1,故障发展区2,预警区,故障区)。图2中分析结果表明,聚类算法能有效的识别机组的运行状态,为设备的维护和检修提供帮助。

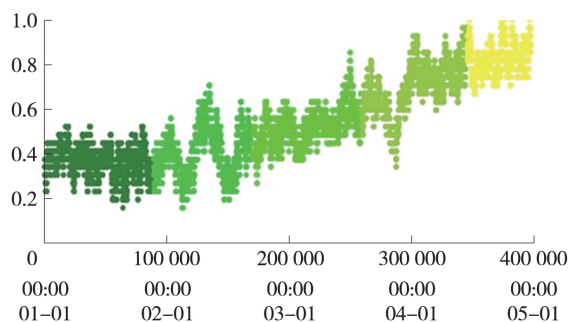


图2 一次设备运行状态聚类分析

### 2.2 基于贝叶斯分类器的状态评估算法

朴素贝叶斯分类器是一系列以假设特征之间强(朴素)独立下运用贝叶斯定理为基础的简单概率分类器。该分类器模型会给问题实例分配用特征值表示的类标签,类标签取自有限集合。它不是训练这种分类器的单一算法,而是一系列基于相同原理的算法:所有朴素贝叶斯分类器都假定样本每个特征与其他特征都不相关。

我们的算法分为3步:

第1步,用采集到的设备状态信息集构建分类器。

第2步,使用构建好的分类器分类未知实例。用来分类的未知实例称作测试实例。一般在分类器被用来预测之前,需要对其分类精度进行评估。只有分类准确率达到要求的分类器才可以用来对测试实例进行分类。

$$y = f(x) = \arg \max_{c_k} \frac{P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)}{\sum_k P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)} \quad (2)$$

第3步,根据当前状态属于不同类别的概率,综合评估机组的当前运行状态。

### 2.3 关联规则分析算法

随着技术的发展,目前已经有很多数据挖掘的算法,其中 Apriori 算法影响最为深远。

后来人们不断对其进行改进。无论如何改进,都有共同的基础定义,即设

$$I = \{i_1, i_2, i_3, \dots, i_m\} \quad (3)$$

是由  $m$  个不同的数据项目组成的集合,其中:元素称为项,项的集合称为项集。给定一个事务数据库

$$D = \{T_1, T_2, T_3, \dots, T_n\} \quad (4)$$

式中:每一个事务  $T$  是项集  $I$  的一个子集,为  $D$  中的总事务数,  $X, Y$  都是  $T$  中的项或项集。如果事务  $T$  同时包含  $X$  和  $Y$ ,那么就可以得到关联规则:

$$X \Rightarrow Y (S\%, C\%, I\%) \quad (5)$$

式中:  $S\%$  为满足条件的事务  $T$  在事务数据库  $D$  中所占比例,即支持度 (Support),

$$S\% = S(X \Rightarrow Y) = P(XY) = |\{T: X \cup Y \subseteq T, T \in D\}| / |D| \quad (6)$$

$C\%$  为  $D$  中包含  $X$  的事务中又包含  $Y$  的比例,即可信度 (Confidence),

$$C\% = C(X \Rightarrow Y) = P(Y|X) = \frac{|\{T: X \cup Y \subseteq T, T \in D\}|}{|\{T: X \subseteq T, T \in D\}|} \quad (7)$$

$I\%$  为  $X$  和  $Y$  的相关程度,即兴趣度 (Interest),

$$I\% = I(X \Rightarrow Y) = \frac{P(Y|X)}{P(Y)} = \frac{|\{T: X \cup Y \subseteq T, T \in D\}|}{|\{T: X \subseteq T, T \in D\}| \cdot |\{T: Y \subseteq T, T \in D\}|} \quad (8)$$

同时大于可信度阈值和支持度阈值的关联规则称为强关联规则。兴趣度大于 1 的强关联规则是有意义的可信规则,便是挖掘的目标。

挖掘问题可分为 2 个子问题: (1) 找出所有满足支持度条件的频繁集; (2) 使用频繁集生成关联规则。其中第 1) 步需多次扫描事务数据库,时间和空间的消耗是制约挖掘效率的关键。频繁模式 (Frequent Pattern, FP) 算法首先将数据库压缩成一棵频繁模式树,相当于将数据库分组,能够减少数据库扫描的次数,而关联信息仍然保存在树的节点中。对于事务数据库  $D$  构造 FP-树的步骤如下:

(1) 扫描  $D$ , 计算频繁项集  $F$  及其支持度,并按支持度计数递减排序,得到频繁项列表  $L$ 。

(2) 创建 FP-树的根结点,对  $D$  中每个事务  $T$  选择  $T$  中的频繁项,并按  $L$  中的次序排序。设排序后的频繁项表为  $F$ ,其中  $F$  是第 1 个元素,而  $P$  是剩余元素的表,如果非空,调用  $\text{insert\_tree}([F|P], T)$  之后如果  $P$  还是非空的,则递归调用  $\text{insert\_tree}(P, N)$ 。

关联规则的挖掘过程分为 3 步:

- (1) 数据预处理;
- (2) 建立挖掘模型;
- (3) 规则抽取和检验。其流程如图 3 所示。

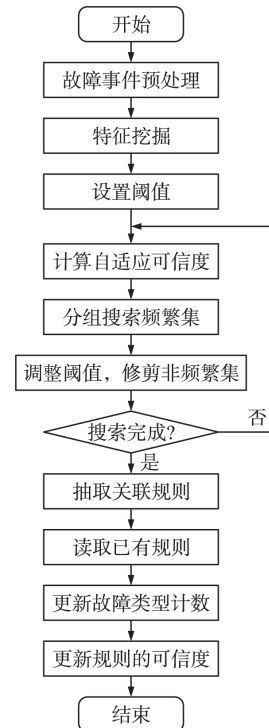


图 3 关联规则挖掘算法流程图

在本算法中,建立挖掘模型就是把预处理之后包含阈值信息的频繁项集构建成 FP-树。该模型包含故障征兆与故障性质之间的关联关系,离线运行时用于故障特征的自学习和故障规则的提取,在线运行时用于故障诊断。故障特征可以通过挖掘故障样本获得,也可以根据已有的运行经验生成。在挖掘系统的运行初期,可以归纳一些故障特征,并在运行过程中不断检验和调整,以适应实际需要。

规则抽取就是选用满足条件的频繁项集,按照与预处理编号过程相逆的步骤生成关联规则。规则检验就是检验新规则是否与原规则库重复或矛盾,以及利用实际的故障记录检验挖掘获得的规则的诊断效果,并与原有规则库的诊断效果进行对比。

### 2.4 状态预测模型

状态预测技术根据机组的结构特点、环境条件和相关历史数据,根据当前运行参数预测、分析和判断机组未来的健康状态,这是实现“诊断后维护”到“预测前维护”的重要手段。

本研究采用大数据分析方法,挖掘出与机组健康状况密切相关的特征量,在此基础上建立了机组振动变化趋势预测模型,为现场设备的运行、维护和改造提供了一定的参考。具体的技术路线是:

(1) 输入变量: 模型的输入变量主要为转速、电流、振动、轴承温度等运行参数;

(2) 模型训练: 首先通过对运行参数的相关性分析挖掘出与振动相关性较大的特征参数, 然后将特征量作为模型的输入完成对预测模型训练;

(3) 振动预测: 模型训练完成后根据机组当前的运行参数预测未来一段时间内振动的变化趋势。

图4给出了某一次设备预测结果。由图中结果可知, 模型预测与实测结果较接近, 即本研究建立的预测模型有较高的预测精度, 能为机组的健康状态评估和劣化趋势判断提供一定的指导依据。

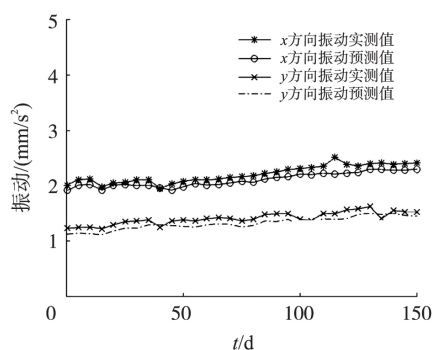


图4 某一次矿井设备状态预测

### 3 系统开发

#### 3.1 系统总体架构

系统总体架构分为5个层级, 分别为设备层、厂级监控层、网络层、集团平台层和远程监控层。

(1) 设备层。主要包括传感器、智能数据采集器、通信设备及附属设施, 通过分布式采集完成设备运行状态的感知、采集与传输, 以实现分布式、网络化的机组运行状态监测网络的构建, 是设备状态评估系统的基础。

(2) 厂级监控层。主要包括机组状态数据服务器、监控分析工程师站、网络通信设施, 完成所有机组运行状态数据的收集、计算、存贮与管理, 实现对

电厂分散机群的集中监控与分析; 整合DCS、SIS等系统的设备运行、管理和状态信息, 构建基于Oracle数据库管理系统的厂级设备运行状态大数据平台; 对振动、机组工况等历史数据进行清洗、标注等预处理, 为多源信息融合、大数据挖掘及机组运行状态评估、劣化分析提供数据准备。

(3) 网络层。用于设备与集团之间的网络通信。

(4) 集团平台层。指建设在集团层面的服务器、大型数据库、存储设备等, 存贮集团各部门所有机组的连续过程状态数据。通过集团数据中心的数据库平台, 构建集团级的机组振动数据库、故障诊断专家知识库和健康评估系统数据库, 为集团层面的监测、诊断、大数据挖掘和智能运维决策提供实时和历史数据服务。

(5) 远程监控层。指基于集团大数据平台上的分析、诊断、预警、报表、统计、决策等各种应用功能模块。在机组振动数据库基础上实现对集团所有机组的监测和分析, 在运行状态评估系统数据库基础上, 实现多源大数据融合、特征提取、仿真计算、建模、机组状态评估及劣化分析等功能; 通过WEB服务和Internet技术, 为网络客户端或远程专家提供面向Internet的数据信息服务和客户端分析服务; 在WEB服务软件基础上, 以多种WEB客户端页面技术, 提供信息发布、面向应用需求的数据浏览、图表分析、数据查询等应用功能。

#### 3.2 状态监测网络

该系统主要由分布式无线数据采集器、工程师站、厂级状态数据服务器、镜像服务器、网络设备、分析软件与远程数据通信等组成, 网络结构如图6所示。

状态监测网络体现了分布式、网络化、独立性、易扩展等特点, 为辅机故障诊断与振动治理提供分析技术手段和数据测试工具。后期可将主机及其他辅机方便地纳入到监测网络, 从而构成网络化的全厂设备故障分析、诊断和决策系统。

##### (1) 分布式无线数据采集器

分布式无线数据采集器基于ARM嵌入式系统, 具备小型化、模块化、网络化、开放性和可扩展性等特点, 既可安装在机柜中用于主机监测, 也能便捷地就地安装监测辅机。采集器将机组状态信号通过WiFi接口发送至全厂状态数据服务器, 可自成一体独立运行, 也可方便组网, 不需要敷设众多的模拟信号电缆和网线, 解决了分散机群的集中监控问题, 为构建分布式网络化全厂监测网奠定了基础。

##### (2) 状态监测工程师站

用于全厂所有数据采集器的数据收集、计算、存

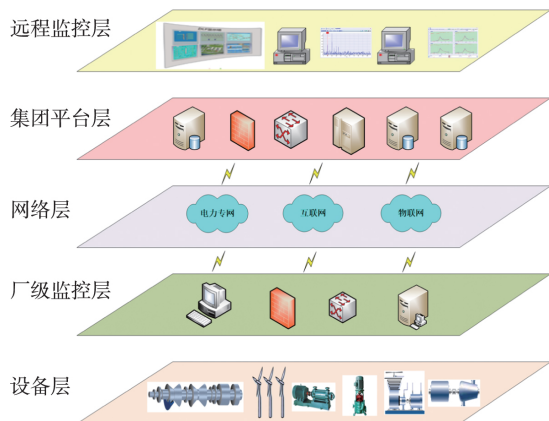


图5 系统总体架构图



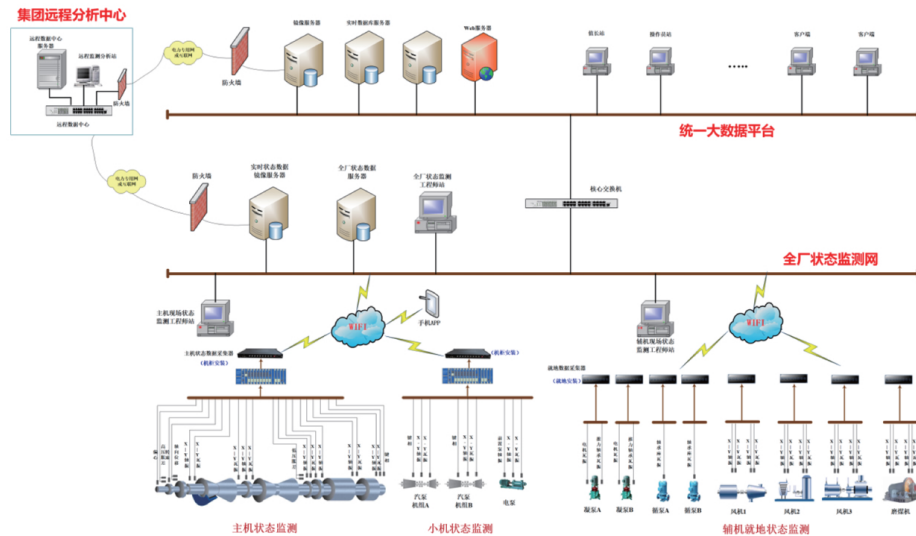


图6 状态监测网络

贮与管理 提供在线监测、振动分析及网络数据服务功能。

### (3) 厂级状态数据服务器

数据来源包括机组振动数据、转速、轴位移、差胀、偏心、缸胀等机组状态数据、来自于 SIS 系统的瓦温、油温、油压、负荷、电量等生产过程参数,以及和机组状态相关的其他数据。基于大数据管理数据库对上述数据进行重构后形成设备运行评估系统的统一数据平台,为状态监测、诊断、评估、预警等提供数据接口及网络通信服务。

### (4) 镜像服务器

振动数据采集频率高、数据量大,由于通信带宽的限制,从设备端发送至集团大数据中心的振动数据是振动倍频幅值与相位等特征值,而非实时振动信号。当远程中心发现故障征兆时,可能还需要获得详细的实时数据以提供更准确的诊断信息。因此,在企业侧配置全厂设备状态数据镜像服务器,通过网闸与外网建立通信服务,当出现振动故障征兆时,远程专家可登录厂内镜像服务器浏览、查询机组的振动实时图谱和历史数据,为准确诊断提供翔实信息,同时也可保证企业数据的安全。

## 5 结束语

以大数据驱动和人工智能算法为基础,开发了矿井旋转机械状态评估成套系统,对矿井旋转机械进行实时故障分析和劣化趋势预测。利用企业系统,融合矿井旋转机械振动、温度、压力、开度、电量等状态参数,通过多源信息融合技术实现矿井旋转机械的智能故障诊断,并对设备的劣化趋势进行预测,为矿井旋转机械的运行管理、维护和预测检修提

供决策依据。同时,构建单机-全厂-集团等多层次数据通信网络架构,为形成涵盖全集团的现场/远程机组智慧运维体系预留空间。该状态评估系统对于提高企业的运维水平和工作效率、探索基于大数据融合的设备诊断与评估方法、促进人工智能与机械行业的深度融合都具有重要意义。

### 参考文献:

- [1] 姚锡凡,刘敏,张剑铭,等. 人工智能视角下的智能制造前世今生与未来[J]. 计算机集成制造系统, 2019, 25(1): 19-34.
- [2] 王万良,张兆娟,高楠,等. 基于人工智能技术的大数据分析方法研究进展[J]. 计算机集成制造系统, 2019, 25(3): 529-547.
- [3] 闪鑫,陆晓,翟明玉,等. 人工智能应用于电网调控的关键技术分析[J]. 电力系统自动化, 2019, 43(1): 49-57.
- [4] 刘云鹏,许自强,李刚. 人工智能驱动的数据分析技术在电力变压器状态检修中的应用综述[J]. 高电压技术, 2019, 45(2): 337-348.
- [5] 樊文飞,王国仁,王朝坤. 智能数据管理与分析技术专刊前言[J]. 软件学报, 2019, 30(3): 495-497.
- [6] Wang Zhiqiang, Liang Jiye, Li Ru, et al. An Approach to Cold-Start Link Prediction: Establishing Connections Between non Topological and Topological Information[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(11): 2857-2870.
- [7] Wu Xindong, Li Yi. Influence Analysis of Online Social Networks[J]. Chinese Journal of Computers, 2014, 37(6): 735-752.



李新虎(1966—),男,陕西眉县人,硕士,陕西能源冯家塔矿业运营有限责任公司总经理, jixinhu\_sn@163.com。