

基于大数据分析的异常检测技术应用于石油催化*

苗永梅

(宝鸡职业技术学院 宝鸡 721304)

摘要: 石油化工领域生产装置异常检测与诊断依赖机理模型和人工经验,存在主观性强、与实际参数不匹配等缺陷。在实践和总结现行石油化工生产装置异常检测与诊断技术的基础上,结合目前主流的大数据分析技术,对生产装置上获得的温度、流量、压力等数据进行实时分析,寻找大数据后面隐藏的规律,让大数据发声,及时预警和处理故障,并做出更明智的决策,减小损失。通过对某催化装置主分馏塔的塔顶温度变化幅度预警测试,比人工提前 6 min 发现故障,将事故消除在萌芽状态。该技术具有工程实用价值,值得在工业控制现场进行实践和推广。

关键词: 大数据;数据挖掘;异常检测;主成分分析

中图分类号: TP391.5 **文献标识码:** A **国家标准学科分类代码:** 510.1050

Anomaly detection based on large data analysis applied to petroleum catalysis

Miao Yongmei

(Baoji Vocational Technology College, Baoji 721304, China)

Abstract: The mechanism and model of anomaly detection and diagnosis for the production equipment in petrochemical industry have the defects of strong subjectivity and mismatch with the actual parameters. Based on the practice and summary of anomaly detection and diagnosis technology of the oil chemical production equipment, combined with the current mainstream data analysis technology, real-time analysis of production equipment on the temperature, flow and pressure data, find the hidden data behind the law, make big data sound, timely warning and fault handling, and to make more informed decisions, reduce losses. Through the test of the tower top temperature variation range of a catalytic converter, it is found that the accident is eliminated in the bud state 6 minutes ahead of time. This technology has practical value in engineering, and it is worthy to be popularized in the field of industrial control.

Keywords: big data; data mining; anomaly detection; principal component analysis

1 引言

石油化工行业是典型的流程型生产模式^[1],工艺过程高度依赖设备的长期、稳定、安全和高效运行,任何一台关键设备出故障都可能给企业造成百万甚至千万级的经济损失。因此,很有必要对设备实时异常检测与诊断^[2],以保证设备长期可靠运行。

石油化工生产装置异常检测与诊断方法有 3 种,基于机理模型^[3]的检测与诊断依赖机理模型,机理模型与工厂实际参数匹配时耦合度不高,导致系统效率不高;基于专家知识与经验^[4]的检测与诊断模型主观性强,缩小了模型的适用程度。

在大数据技术的支撑下,用数据分析、数据挖掘技术^[5]

对工业生产过程中设备运行的状态数据分析,提前捕捉设备的故障征兆,及时采取补救措施,减小损失。本文以炼油厂的催化装置数据为例,进行数据分析和运算,阐述基于大数据分析的实时异常检测与诊断技术给石化企业安全生产带来的变化与价值。

2 数据源分析

从石油化工生产装置的控制中直接采集的数据(轴承振动、温度、压力、流量),数据量大、输入和处理速度快、数据多样性、价值密度低^[6],这些非结构化的数据无法用传统

的数据处理工具来分析和处理。如何管理这些数据,并从这些数据中搜寻有用的信息,为人类更好的服务,成为

收稿日期:2017-03

* 基金项目:国家自然科学基金(51405382)、省级课题“互联网+”创新教育对策研究(SZJG-1629)资助项目

人们研究这些大数据的驱动因素。

大量的数据只有经过分类、加工、整理、分析才能满足不同的要求。而要处理大量的数据,就需要大数据技术的支持。用数据挖掘技术寻找其关联性,用分类回归^[7]、神经网络^[8]、自适应算法^[9],对所采集到的数据进行分析、测试,确定报警阈值,让大数据发声,帮助企业适应变化,并做出更明智的决策。

3 数据处理方法

源数据又称脏数据,干扰多,误差大^[10],需要进行筛选和计算。工业实时数据进入模型计算,首先需要进行稳态识别^[11]和归一化处理^[12],如果数据缺失较为严重,还需要进行缺失值^[13]的填充。

数据处理方法采用主成分分析^[14-15]。主成分分析(PCA)是一种分析、简化数据集的技术,通过保留低阶主成分忽略高阶主成分,减少数据集的维数,低阶成分能够保留住数据的最重要方面。

数据的主成份用 F_1 (选取的第一个线性组合,即第一个综合指标)的方差来表达,即 $\text{var}(F_1)$ 越大,表示 F_1 包含的信息越多。因此在所有的线性组合中选取的 F_1 应该是方差最大的,故称 F_1 为第一主成分。如果第1主成分不足以代表原来 P 个指标的信息,再考虑选取 F_2 (即第2个线性组合),为了有效地反映原来信息, F_1 已有的信息就不需要再出现在 F_2 中,即 F_1 、 F_2 的相关系数为0,用公式表示为 $\text{cov}(F_1, F_2) = 0$,依此类推可以构造出第3,第4, ..., 第 F_p 个主成分,其公式表示如下:

$$F_p = a_{1i} \cdot ZX_1 + a_{2i} \cdot ZX_2 + \dots + a_{pi} \cdot ZX_p$$

其中 a_{1i} 、 a_{2i} 、 a_{pi} ($i = 1, \dots, m$)为 X 的协方差阵 Σ 的特征值所对应的特征向量, ZX_1 、 ZX_2 、 ZX_p 是原始变量经过标准化处理的值。

这种运算可以被看作是揭露数据的内部结构,从而更好的解释数据变量的方法。如果一个多元数据集能够在高维数据空间坐标系中被显现出来,那么PCA就能够提供一幅比较低维度的图像,这幅图像即为在讯息最多的点上对原对象的一个‘投影’,去掉非主成份降低数据的维度。

4 数据预警

对生产现场得到的大数据预处理后,可进行加工处理,确定故障阈值,超过阈值则报警,以便及时采取补救措施,减小损失。

4.1 幅值预警

幅度变化是异常判断的一个重要指标,在专家知识中使用频率非常高。可以通过温度变化幅度、速度变化幅度、压力变化幅度、气体扩散幅度等因素来分析和监测异常。幅度变化预警的难点在于用于计算的基准是相对的,既不能一成不变,也不能频繁变化,否则会导致计算结果出现偏差。

变化幅度计算的数据,需要采集基于一段时间的历史数据,用运动的时间窗来提取最新的数据,并对这个时间窗里的数据进行筛选,舍弃非稳态的数据。变化幅度计算的另一个重要注意事项是,不能采用一次数据进行直接计算,而是需要进行平滑处理,否则计算结果波动过大,失去参考的价值。图1所示为一次数据直接计算和数据回归后进行计算的不同变化幅度曲线,显示回归计算结果更有参考价值。

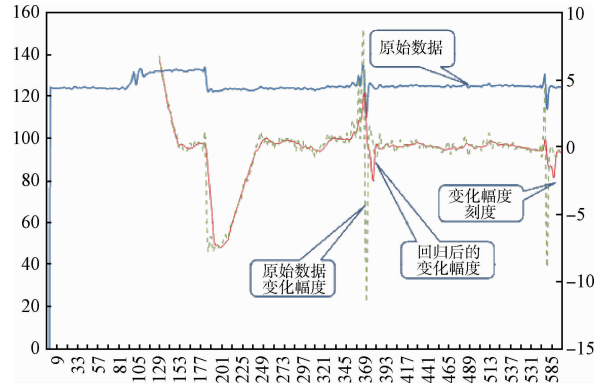


图1 直接与回归计算幅度曲线

比较典型的应用是监测升降速率。当升降速率变化时,在触发报警之前,不太容易被人观察到,但通过对采集的数据实时分析,提前发现异常。图2所示为某催化装置主分馏塔的塔顶温度,如设定温度升降速率报警,在此次事故发生前6 min,操作或管理人员就收到了警报,则可将事故消除在萌芽状态,最大限度地避免事故扩大。

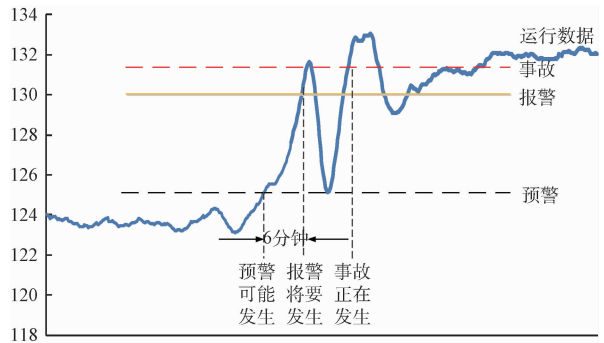


图2 温度升降速率预警示

4.2 多变量预警

从石油化工生产装置上获得的温度、流量、压力等数据,用某一项数据(温度变化幅度)分析预警,称为单变量预警;单变量预警的特点是目标明确,针对性强,能够解决已知异常或事故的监测。但针对未知异常,单变量预警就显得无能为力了,需要考虑多变量的综合因素。用多个单变量综合分析组成一个事故预测模型,用于实时监测,图3所示为某催化装置的外取热爆管事故监测模型,用到了料位

的归零计算,压力的升降速率计算,以及时间参数。

基本参数				
模型名称: 外取热爆管或泄漏预警		所属模板: 外取热爆管或泄漏现象预警		
参数定义				
参数名	类型	位号名	参数描述	
<input type="checkbox"/>	时间窗	输入参数	20	参数设置
<input type="checkbox"/>	外取热料位升降速率	输入参数	YC_LI6104.RATE	参数设置
<input type="checkbox"/>	外取热料位升降速率设定值	输入参数	-0.1	参数设置
<input type="checkbox"/>	外取热料位归零	输入参数	YC_LI6104.ZERO	参数设置
<input type="checkbox"/>	外取热器压力升降速率	输入参数	YC_PI6113.RATE	参数设置
<input type="checkbox"/>	外取热器压力升降速率设定值	输入参数	0.00001	参数设置
<input checked="" type="checkbox"/>	事故可能发生	输出参数	PAR_0106	参数设置

图 3 多变量组合检测示例

多变量异常检测与诊断需要分两步进行:1)构建明确的物理关系,即将所有的工艺参数,按物理和行政结构构建层次关系,让每一个参数都隶属一个设备和操作单元。如催化装置反应沉降器的提升管出口温度、压力、料位同属于一个设备,存在着必然的内在联系。2)引入大规模数据计算和扫描,将异常检测与诊断的范围扩大到设备、操作单元以及系统和装置层面,确保每个操作参数都受到监测。

4.3 主成份分析预警

主成分分析用于异常检测与诊断的方法如下:观察当前工况在主元上的“得分”趋势,即可用这一新的单一统计变量监控生产过程的异常。通常只需针对第一主元得分进行趋势监控;复杂工况需对 2 个主元进行趋势监控即可得到满意的效果。

对归一化后的变量进行主元分析后,主元得分的结果基本符合正态分布的假设,可使用简单的判读,如 99.7% 的可能性应在在 $[-3, 3]$ 区间内,绝对值超出 3 基本上可判定为异常,图 4 所示为某炼油厂催化装置外取热器爆管时的相关参数的主成分分析结果,当异常出现时,主元得分明显偏离正常分析分析结果,当异常出现时,主元得分明显偏离正常值。可见,在物理关系约束下,主成分分析能够很好

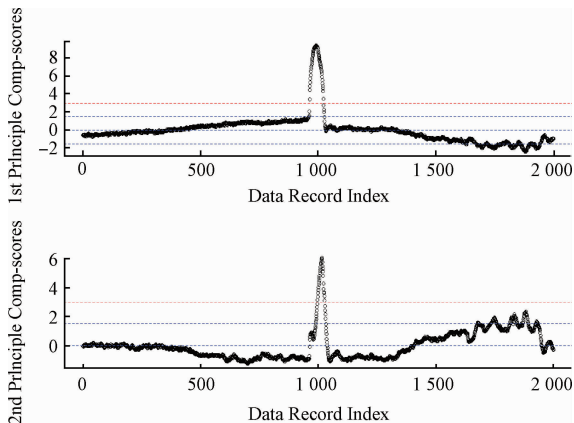


图 4 主成分分析诊断结果

地弥补单变量检测覆盖不全的缺点,两者结合后,能够实现从点到面的全面覆盖。

5 工程进步

目前报警管理的一个难点就是如何减少无效报警量,无效报警可能是偶发,也可能是参数设置不合理。基于大数据的实时异常检测与诊断系统,超限报警仅仅是其中一种方式,而更多的是采用连续的数据分析进行跟踪和报警,这种方式能够最大限度地消除偶发报警,当偶发报警出现时,由于有更多的冗余信息进行判断,便于操作和管理人员识别报警的真实性和严重性。

当系统为管理层服务时,需要具备这样一个功能,即管理人员关注极少数重要的参数,以及报警情况,但也不想遗漏突发事件,此时多变量异常检测就发挥作用了。由于多变量异常检测以操作单元为对象,故该操作单元及内部设备出现的异常,都将被检测出来,及时将报警信息送到管理人员的移动设备上。

6 结 论

实现工作现场的实时报警,得益于大数据技术的支持,采集和分析历史数据,及时纠正,使系统的整体性能得到提高,为企业提供高效的信息化服务。

经过研发、测试和现场实施,同时具备单变量和多变量异常检测与诊断的名道恒通工厂异常实时监管软件(PARS)已成为石化企业安全生产的管理利器,在济南炼化的实际应用中,取得了良好的效果,帮助用户避免了公用工程、设备、操作等方面因异常处理不及时而导致的事态扩大和经济损失,能够为企业的安全生产提供更有力的技术保障。

参考文献

- [1] 王维彤. ERP 系统在石化行业之应用[J]. 中国科技信息 2005(14):121-121. <http://news.163.com/15/0715/16/AUJ2R77A00014AED.html>
- [2] 严英杰,盛戈璋,陈玉峰,等. 基于大数据分析的输变电设备状态数据异常检测方法[J]. 中国电机工程学报, 2015(2): 45-46.
- [3] 肖海林,彭星闻,王方华. 企业持续发展的生成机理模型:基于海尔案例的分析[J]. 管理世界, 2004(8): 111-118.
- [4] 卞玉涛,李志华. 基于专家系统的故障诊断方法的研究与改进[J]. 电子设计工程, 2013, 21(16):83-86.
- [5] 彭宇,庞景月,刘大同. 大数据:内涵、技术体系与展望[J]. 电子测量与仪器学报, 2015, 29(4):469-482
- [6] 维克托·尔耶·舍恩伯格,肯尼斯·库克耶. 大数据时代:生活、工作与思维的大变[M]. 杭州:浙江人民出版社, 2013.

(下转第 187 页)