

Fault Diagnosis of Screw Pump Based on Correlation and Multi-Order Fitting

Guanglian Lin¹, Xiuli Shao¹, Qiang Wu²

¹College of Computer Science, Nankai University, Tianjin

²Tianjin Pump Industry Machinery Group Co., Ltd., Tianjin

Email: linguanglian@mail.nankai.edu.cn, shaoxl@nankai.edu.cn, yanfashi@ctppumps.com

Received: Feb. 24th, 2020; accepted: Apr. 6th, 2020; published: Apr. 13th, 2020

Abstract

The vibration acceleration of the screw pump at different center frequencies deviates from the normal range, which will cause abnormal vibration of the screw pump. Therefore, the screw pump needs to set different measurement points to detect its vibration acceleration before leaving the factory. For this reason, based on the vibration acceleration of different measurement points of the screw pump under normal operation, this paper applies the Apriori algorithm to obtain the conclusion that there is no significant difference in vibration acceleration between different measurement points of the screw pump and that there is a strong positive correlation. The multi-order fitting method is used to establish the functional relationship between the center frequency and the vibration acceleration of the screw pump. The coefficient vector corresponding to the normal operating parameters is used as a template, and calculate the coefficient distance between the coefficient vector constructed by the operating parameters of the measurement points of the equipment to be diagnosed and the template, then apply statistical methods to design an appropriate threshold to compare with it, and determine the sample to be diagnosed as a fault if and only if the coefficient distance is bigger than the threshold. After practical application, the method in this paper can find the abnormal vibration of the screw pump in time, so as to achieve the purpose of fault diagnosis.

Keywords

Apriori Algorithm, K-Means, Least Squares, Multi-Order Fitting, Fault Diagnosis

基于关联及多阶拟合的螺杆泵故障诊断

林广莲¹, 邵秀丽¹, 武强²

¹南开大学计算机学院, 天津

²天津泵业机械集团有限公司, 天津

Email: linguanglian@mail.nankai.edu.cn, shaoxl@nankai.edu.cn, yanfashi@ctppumps.com

收稿日期：2020年2月24日；录用日期：2020年4月6日；发布日期：2020年4月13日

摘要

螺杆泵在不同中心频率下的振动加速度偏离正常范围会导致螺杆泵产生异常振动，因此，螺杆泵在出厂前需要设置不同测点检测其振动加速度。为此，本文基于螺杆泵在正常工作下不同测点的振动加速度应用Apriori算法得到螺杆泵不同测点间振动加速度并无显著性差异且呈强正相关的结论；然后应用多阶拟合法建立螺杆泵中心频率和振动加速度的函数关系式，以正常运行参数所对应的系数向量作为模板，计算出待诊断设备的测点运行参数所构造的系数向量与模板之间的系数距离，并应用统计方法设计适当的阈值与之作比较，当且仅当系数距离大于阈值时，将待诊断样本判定为故障。经实际应用，本文方法能够及时发现螺杆泵的异常振动，从而达到故障诊断的目的。

关键词

Apriori算法，K-Means聚类，最小二乘法，多阶拟合，故障诊断

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

螺杆泵作为一种排水及动力装置，因其具有平衡性好、启动时间短和振动噪声小等优点，常常作为辅机，装备在船舶、航母及潜艇等大型军事设备中。因此，螺杆泵能正常、安全、高效地运行不仅关系到泵器件本身，更关乎舰船的运行安全[1]。不仅如此，诸如潜艇此类军事设备，为了保证作战时的隐蔽性，必须要求舰艇内部设备及机械装置的噪声、振动、声波严格控制在安全范围中。螺杆泵的异常振动会极大地影响舰艇的隐蔽性，从而对整个舰艇和舰艇上的人员的安全造成威胁。因此，维持螺杆泵的正常运行，检测螺杆泵存在的异常振动具有实际意义，基于螺杆泵的振动指标对螺杆泵进行异常检测成为近些年的研究热点。

之前已经有许多学者提出了异常检测所存在的问题和解决方法。由于采集设备故障数据的成本较高且难度较大，因此大部分的实验数据都是在设备正常工作下检测得到的，Zaher 等人[2]将缺乏故障记录描述为早期故障检测的主要挑战，从而强调了正常行为建模的重要性。Shen Zhang 等人[3]提出了一种使用基于变分自动编码器(VAE)的深度生成模型进行轴承异常检测的半监督学习方法，当只有一小部分数据带有标签时，该方法可以有效利用数据集。Sreelekha Guggilam 等人[4]提供了一种流式聚类和异常检测算法，该算法在执行概率异常检测和聚类时不需要对异常分数或簇数有严格的阈值，这确保了群集的形成不会受到异常数据的存在的影响，从而能够更可靠地定义“正常与异常”行为。

针对上述背景，本文基于螺杆泵在正常工作下的特征参数(不同中心频率下的振动加速度)，深入研究螺杆泵的振动异常检测方法。首先根据企业方提供的螺杆泵运行数据应用 K-Means 进行离散化处理，然后利用 Apriori 算法深入挖掘正常工作下不同测点振动加速度间的潜在关系。并在此基础上，基于螺杆泵中心频率和振动加速度的函数关系应用多阶拟合法，求解多项式拟合系数，以正常运行参数所对应的系数向量作为模板，计算出待诊断设备的测点运行参数所构造的系数向量与模板之间的系数距离，根据最

小距离原则，距离越小表明待诊断样本与模板越相似。为进一步量化诊断规则，本文通过统计方法，设计并计算适当的阈值，将其与系数距离作比较，从而得出故障诊断结果。最后，为了验证方法的有效性和可行性，本文通过模拟仿真试验，收集故障数据并构建测试数据集，对测试数据集中的样本应用多阶拟合法并计算相应的系数距离，根据阈值得出诊断结果。另外，计算相应的评估指标(正确率、精度及召回率等)，从而对实验结果进行分析。

2. 螺杆泵实验数据

本文分析的数据源是企业对正常工作的泵进行检测所得到的实验数据。针对同一型号的泵，按 GJB4058-2000 《J 船设备噪声、振动测量方法》中的相关规定，对机脚振动加速度进行振动测量，泵组机脚振动加速度测点位置如下图 1 所示。

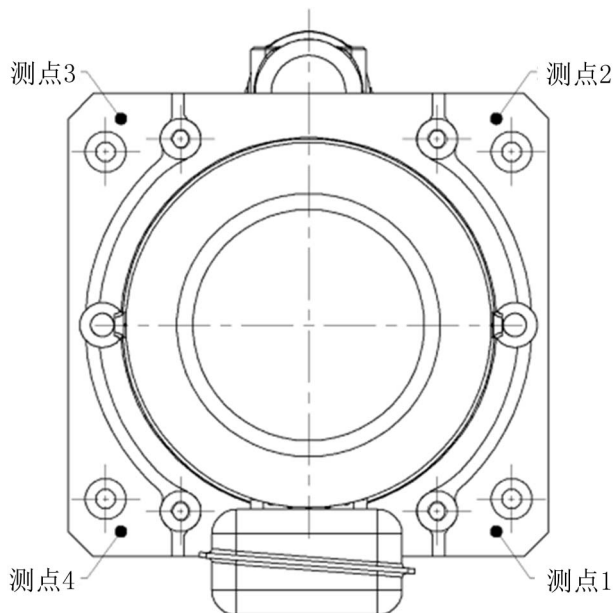


Figure 1. Location of vibration acceleration measurement point of pump set feet

图 1. 泵组机脚振动加速度测点位置

企业方已经应用傅里叶变换将测量数据从时域变换到频域，因此，本文的数据源字段主要包括中心频率和各个测点的振动加速度，描述的是不同中心频率所对应的各个测点的振动加速度，展示部分数据参见表 1。

Table 1. Some data sources

表 1. 部分数据源

中心频率	测点 1	测点 2	测点 3	测点 4
10	65.4	66.7	66.8	66.9
12.5	72	72.1	71.5	73.9
16	72.8	71.9	70.7	75
20	72.8	72.3	73.6	74.3
25	90.6	90.1	81.5	83.5

3. 基于 Apriori 算法的关联规则分析

由表 1 可以看出, 测点之间的差异并不显著, 因此本文进一步挖掘各测点之间是否存在互相影响的关系。首先应用 K-Means 对实验数据进行离散化处理, 然后基于 Apriori 算法[5]和相关系数探究测点之间的潜在关系。一方面能够避免相似测点对诊断结果的影响, 另一方面能够简化后续模型求解过程。

3.1. 基于 K-Means 的数据离散化

由于 Apriori 算法要求离散的数据格式, 而实验数据是连续数值型数据, 故需要先对数据进行离散化处理。本文采用基于 K-means 的一维聚类离散化方法, 先确定最佳 K 值, 再将连续属性值划分为 K 个区间, 让簇内的点差值小, 而让簇间点差值大。

本文采用肘部法则(Elbow Method) [6]确定 K 值。此方法适用于 K 值相对较小的情况, 将每个簇的质点与簇内样本点的平方距离误差和称为畸变程度(Distortion), 当选择的 k 值小于真正的 K 时, k 每增加 1, Distortion 值就会大幅的减小; 当选择的 k 值大于真正的 K 时, k 每增加 1, Distortion 值的变化就不会那么明显。这样, 正确的 K 值就会在这个转折点, 类似手肘的地方。

如图 2 所示, 通过绘制 k 与 Distortion 的关系曲线图确定最佳 K 值, 肘部的值(Distortion 开始时下降很快, 在肘部开始平缓了)作为 K 值, 可以看出 K 应取 3。对每个测点都应用肘部法求 K 值, 得到各个测点的最优 K 值均为 3。

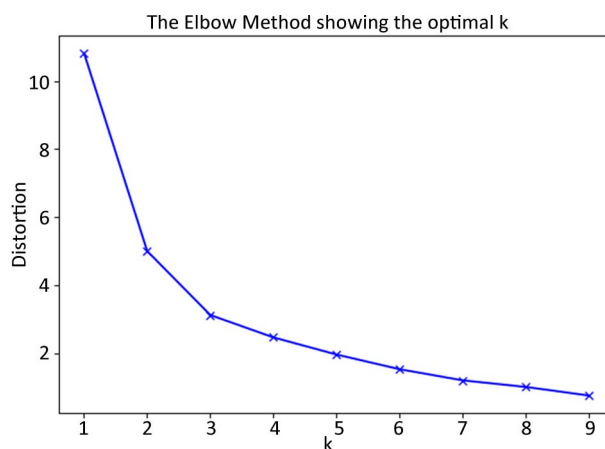


Figure 2. Curve: applying measurement point 1 to the elbow method
图 2. 测点 1 应用肘部法

根据每个测点的最优 K 值, 应用 K-means 对各个测点数据进行聚类, 从而达到离散化的目的。K-means 的基本思想是先从样本集中随机选取 K 个样本作为簇中心, 并计算所有样本与这 K 个“簇中心”的距离, 对于每一个样本, 将其划分到与其距离最近的“簇中心”所在的簇中, 对于新的簇计算各个簇的新的“簇中心”。直到“簇中心”不再变化或达到 300 次迭代为止。

将离散化结果中的每个测点数据标记为 K 个振动加速度级, 例如测点 1 的 K 值为 3, 则将其离散化结果分别标记为“一级”, “二级”和“三级”。

3.2. 关联规则分析

为挖掘不同测点之间的关系, 基于离散化后的数据, 本文采用 Apriori 算法分析不同测点的振动加速度级之间的关联规则。该算法的基本思想是: 首先利用递归的方法[7]找出所有的频繁项集(诸如{测点 1 = “二级”, 测点 2 = “二级”}等), 然后由频繁项集产生强关联规则(诸如测点 2 = “二级” \rightarrow 测点 1 = “二

级”等), 将大于给定最小置信度的规则保留下来, 具体的算法流程如图 3 所示。

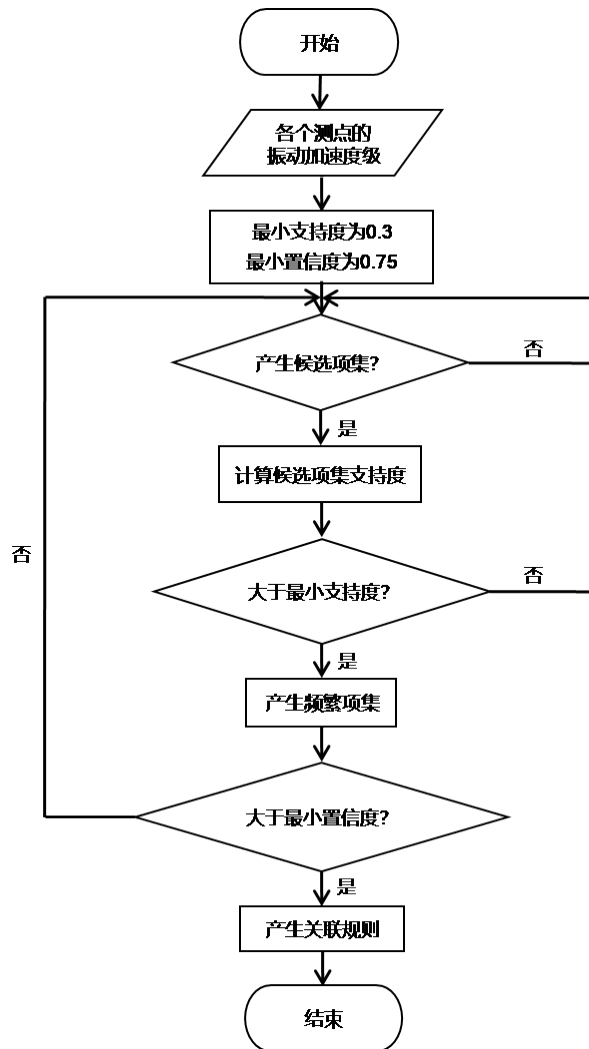


Figure 3. Flow chart of association rule analysis based on Apriori
图 3. 基于 Apriori 的关联规则分析流程图

基于 Apriori 的关联规则分析流程图如图 3 所示, 将最小支持度和最小置信度分别设置为 0.3 和 0.75, 总共得到 48 条关联规则, 单属性对的关联关系共 12 条, 多属性对之间的关联关系共 36 条。观察得知多属性组之间的关联关系均依托于单属性对之间的关联关系, 数量较多, 故本文仅就部分单属性关联规则结果进行展示, 参见表 2。

Table 2. Association rules
表 2. 关联规则

关联规则	置信度
测点 3 = “二级” → 测点 2 = “二级”	0.91
测点 2 = “二级” → 测点 4 = “二级”	0.92
测点 4 = “二级” → 测点 1 = “二级”	0.92
测点 1 = “二级” → 测点 3 = “二级”	0.79

由表 2 可以发现, 测点之间的关联规则能够形成一个循环, 即: 测点 3 = “二级” → 测点 2 = “二级” → 测点 4 = “二级” 测点 1 = “二级” → 测点 3 = “二级”。因此, 对于正常工作下的螺杆泵, 在不同的中心频率下, 不同测点的相同振动加速度级总是同时出现的, 即不同测点间的振动加速度并无显著性差异。

3.3. 基于相关系数的不同测点间的相关性分析

为验证不同测点间的相关性, 本文考虑基于相关系数进行相关性分析。相关系数是用以反映变量之间相关关系密切程度的统计指标[8]。由于研究对象的不同, 相关系数有多种定义方式, 当研究数据呈正态分布时较多采用皮尔逊相关系数, 不满足正态分布时较多使用斯皮尔曼相关系数。本研究中的数据不满足正态分布, 故使用斯皮尔曼相关系数。

斯皮尔曼相关系数[9]定义如下: 对于样本容量为 n 的样本, n 个原始数据被转换成等级数据, 斯皮尔曼相关系数 ρ 为

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

x_i, y_i 代表降序后的等级数据, \bar{x}, \bar{y} 分别代表 x 和 y 的均值。

计算测点间的相关系数, 结果如表 3 所示。可以发现, 测点间的相关系数均在 0.92 以上, 呈强正相关, 印证了由关联分析得到的不同测点间并无显著性差异的结论。

Table 3. Correlation coefficient between each measurement point
表 3. 各测点间的相关系数

	测点 1	测点 2	测点 3	测点 4
测点 1	1	0.921425	0.952137	0.95748
测点 2	0.921425	1	0.94047	0.9367976
测点 3	0.952137	0.94047	1	0.9462497
测点 4	0.95748	0.936798	0.94625	1

4. 多阶拟合法

由上文的分析可知, 测点间并无显著性差异, 因此单测点分析和多测点分析并不对诊断结果造成影响。在此基础上, 应用多阶拟合法建立中心频率和振动加速度的函数关系式, 以系数向量间的欧式距离作为诊断依据, 将其与适当的阈值作比较从而得出故障诊断结果。

4.1. 模型的建立

在螺杆泵正常运行的前提下, 测点间无显著性差异, 因此可用任一测点代表其余测点。本文以中心频率为自变量, 任一测点作为因变量求解数学模型, 函数关系式形式如下, 其中 x 表示中心频率, ω_i 表示各阶的权重。

$$f(x) = \omega_0 + \omega_1 x + \dots + \omega_n x^n = \sum_{i=0}^n \omega_i x^i$$

多项式拟合是曲线拟合方法之一, 目的是使求得的拟合曲线能够反应数据的总体分布, 从而使数据能够分布在拟合曲线的附近。为了评价拟合函数的优劣, 需要建立损失函数, 测量每个样本点目标值

与预测值之间的误差，拟合的目标是让误差最小。

假设有数据点 $((x_1, y_1), (x_2, y_2), \dots, (x_m, y_m))$ ，其中 x_i 表示中心频率， y_i 表示振动加速度。本文应用最小二乘法[10]，求解拟合函数解析式，均方误差函数为

$$\delta(\omega_0, \dots, \omega_n) = \sum_{i=1}^m [f(x) - y_i]^2$$

求解出最小 δ 值，即可解出 $f(x)$ 的函数系数 $\omega_1, \omega_2, \dots, \omega_n$ 。为求得误差函数的最小值，对函数两边求导。

$$\begin{aligned} \frac{\partial \delta}{\partial \omega_0} &= 2 \sum_{i=1}^m [f(x) - y_i], \\ \frac{\partial \delta}{\partial \omega_1} &= 2 \sum_{i=1}^m [f(x) - y_i] x, \\ &\vdots \\ \frac{\partial \delta}{\partial \omega_n} &= 2 \sum_{i=1}^m [f(x) - y_i] x^n \end{aligned}$$

令上述等式为零，得到如下正规方程组：

$$\begin{cases} m\omega_0 + \sum_{i=1}^m \omega_1 x_i + \dots + \sum_{i=1}^m \omega_n x_i^n = \sum_{i=1}^m y_i \\ \sum_{i=1}^m \omega_0 x_i + \sum_{i=1}^m \omega_1 x_i^2 + \dots + \sum_{i=1}^m \omega_n x_i^{n+1} = \sum_{i=1}^m x_i y_i \\ \vdots \\ \sum_{i=1}^m \omega_0 x_i^n + \sum_{i=1}^m \omega_1 x_i^{n+1} + \dots + \sum_{i=1}^m \omega_n x_i^{2n} = \sum_{i=1}^m x_i^n y_i \end{cases}$$

将上式改写成矩阵形式：

$$\begin{bmatrix} m & \sum_{i=1}^m x_i & \dots & \sum_{i=1}^m x_i^n \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 & \dots & \sum_{i=1}^m x_i^{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m x_i^n & \sum_{i=1}^m x_i^{n+1} & \dots & \sum_{i=1}^m x_i^{2n} \end{bmatrix} \begin{bmatrix} \omega_0 \\ \omega_1 \\ \vdots \\ \omega_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \\ \vdots \\ \sum_{i=1}^m x_i^n y_i \end{bmatrix}$$

可以证明，由于方程组的系数矩阵是一个对称正定矩阵，因此存在唯一解 $\omega_1, \omega_2, \dots, \omega_n$ 。

为探究在本实验中函数拟合的最佳阶数 n ，本文利用测点 1 的数据，拟合 1 至 9 阶的多阶函数进行对比分析。不同阶数所对应的均方误差及 R^2 如表 4 所示，均方误差回归损失越小且 R^2 越接近 1，则说明拟合效果越好，从而确定最优的“最大阶数”。

Table 4. Mean square error and R^2 of different orders
表 4. 不同阶数的均方误差及 R^2

阶数	均方误差(RMSE)	R^2
1	12.04	0.23
2	9.6	0.51

Continued

3	7.41	0.71
4	6.29	0.79
5	8.97	0.57
6	11.04	0.35
7	12.54	0.16
8	13.24	0.06
9	13.59	0.01

由表 4 可知, 阶数为 4 所对应的 RMSE 最小, R^2 非负且接近于 1, 因此所对应的模型最佳。为了更直观的展示拟合效果, 图 4 是在不同阶数下, 拟合中心频率和测点 1 的关系曲线示意图, 可以发现, 四阶及以上拟合程度较高。当设置阶数越高, 震荡越明显, 出现过度拟合现象。因此, 阶数为 4 的拟合函数的确能够得到最佳模型。在阶数为 4 时, 可得到如下函数模型:

$$y = -2.532 \times 10^{-13} \cdot x^4 + 4.304 \times 10^{-9} \cdot x^3 - 2.37 \times 10^{-5} \cdot x^2 + 0.04688 \cdot x + 77.41$$

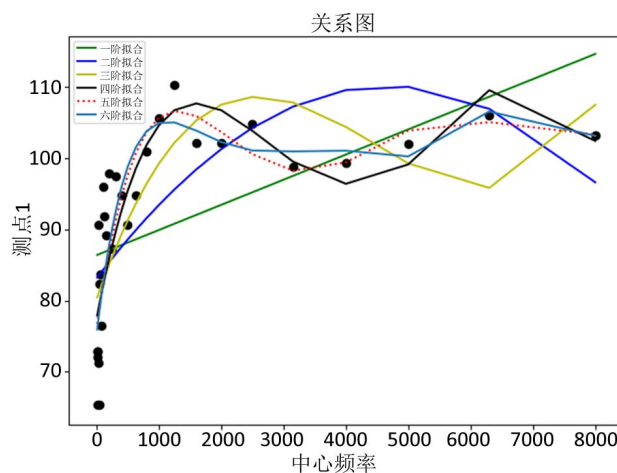


Figure 4. Fitting curves of center frequency and measurement Point 1 under different orders

图 4. 中心频率和测点 1 在不同阶数下的拟合曲线

4.2. 阈值的确定

由 4.1 节可知, 阶数为 4 的拟合函数最优。针对正常工作下的数据集, 建立多个 4 阶一元回归模型, 以中心频率为自变量, 每个测点的振动加速度为因变量, 相应的得到多个拟合系数向量。计算每个拟合系数向量的阶数相同的各个分量的标准差 $\sigma_i, i=1,2,3,4$ 。记 X 为测点拟合系数的数字矩阵, 则

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{bmatrix}$$

其中 n 代表测点数, 列数为 4 对应模型的最佳拟合阶数, 也就是说为每个测点都建立了一个一元 4 次回归模型, 每一行对应每个测点的拟合系数向量。各个测点所对应的拟合系数向量参见表 5。

Table 5. Fitting coefficient vector of each measurement point
表 5. 各个测点的拟合系数向量

	系数向量
测点 1	$(-2.532 \times 10^{-13}, 4.304 \times 10^{-9}, -2.37 \times 10^{-5}, 0.04688)$
测点 2	$(-2.655 \times 10^{-13}, 4.552 \times 10^{-9}, -2.524 \times 10^{-5}, 0.04967)$
测点 3	$(-2.244 \times 10^{-13}, 3.95 \times 10^{-9}, -2.248 \times 10^{-5}, 0.04566)$
测点 4	$(-2.334 \times 10^{-13}, 4.065 \times 10^{-9}, -2.301 \times 10^{-5}, 0.04688)$

根据拟合系数向量求解 $\sigma_i, i=1,2,3,4$ 的计算公式如下, 其中 \bar{x}_i 为第 i 列的平均值。

$$\sigma_i = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2}{n-1}$$

最后根据 3σ 原则, 阈值大小由一个正常分布在 99.73% 概率的置信区间的 3σ 正态分布值来决定, 具体计算公式如下。

$$\theta = \|3\sigma_1, 3\sigma_2, 3\sigma_3, 3\sigma_4\|_2$$

根据公式计算得到阈值大小为 0.0051。为利用上述阈值进行故障诊断, 应用多项式拟合法得到任一正常工作下测点的拟合系数向量 \mathbf{x} 与待诊断设备的测点所对应的拟合系数向量 \mathbf{y} , 二者之间的欧式距离 $Distance$ 与阈值作比较, 大于阈值即为故障。

$$Distance = \sqrt{(\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y})}$$

因此, 判断故障就可以通过评估 $Distance$ 的大小是否超过一个代表正常工作下测点所对应的拟合系数向量的可变化程度的阈值。

5. 基于多阶拟合法的故障诊断

为了验证多阶拟合法行之有效, 本文通过仿真实验, 模拟泵在故障时的运行状态, 收集故障测点数据。将正常数据和故障数据构成测试数据集, 并应用上述方法进行实验。由于不同测点近乎等价, 因此可以选用正常工作下的任一测点作为模板数据, 模板数据所对应的拟合系数向量为模板系数向量, 分别计算模板系数向量与测试数据集中所有数据样本的系数向量的欧氏距离, 部分结果如表 6 所示。

Table 6. Calculation results of template data and test set data
表 6. 模板数据与测试集数据的计算结果

系数向量	距离	实际标签	预测标签
$(-2.66E-13, 4.55E-09, -2.52E-05, 4.97E-02)$	0.00278971	正常	正常
$(-2.24E-13, 3.95E-09, -2.25E-05, 4.57E-02)$	0.00121811	正常	正常
$(-2.33E-13, 4.07E-09, -2.30E-05, 4.69E-02)$	1.07E-06	正常	正常
$(-4.98E-13, 8.22E-09, -4.34E-05, 8.83E-02)$	0.00533182	正常	故障
$(-3.91E-13, 4.01E-09, -8.95E-06, 2.75E-03)$	0.04412941	故障	故障
$(-1.61E-12, 2.24E-08, -9.00E-05, 1.03E-01)$	0.05620737	故障	故障
$(2.95E-13, -3.92E-09, 1.41E-05, -1.92E-02)$	0.06605891	故障	故障
$(5.53E-13, -7.07E-09, 2.51E-05, -2.69E-02)$	0.07382046	故障	故障
$(3.29E-13, -4.93E-09, 1.83E-05, -1.43E-02)$	0.06114535	故障	故障

当且仅当系数距离不超过 4.2 节中得到的阈值 0.0051 时, 才将测点判定为正常, 反之为故障。由上表可以看出, 多阶拟合法能够诊断出所有的故障测点, 但是会出现将正常测点划分为故障测点的情况, 这很有可能意味着该测点所检测的螺杆泵存在潜在的故障因素, 如设备运行时间过长从而导致内部零件受损等。由于将故障样本判定出错的代价远高于将正常样本判定错误的代价, 因此, 本文所提出的多阶拟合法及阈值的确定方法能够较为准确的对螺杆泵进行异常检测, 将判定结果以混淆矩阵的形式展现, 参见表 7。

Table 7. Confusion matrix of fault diagnosis by multi-order fitting method

表 7. 应用多阶拟合法进行故障诊断的混淆矩阵

	正常测点	故障测点
正常测点	41	9
故障测点	1	49

基于混淆矩阵计算的评价指标参见表 8, 基于多阶拟合法的故障诊断的准确度高达 0.9, 就 F 值(0.89)而言, 查全率和精确度之间达到了良好的平衡。因此, 本文所提出的多阶拟合法具有良好的应用前景, 具有实际意义。

Table 8. Evaluation index of multi-order fitting

表 8. 多阶拟合法的评价指标

准确度	查全率	精度	F-Score
0.9	0.98	0.82	0.89

6. 结束语

本文的数据源是基于螺杆泵在正常工作下得到的, 为挖掘数据之间的关系, 本文通过 Apriori 算法发现螺杆泵正常工作下不同测点振动加速度间并无显著性差异, 甚至近乎等价, 且呈现出强正相关。因此, 尽管数据项多, 但是单属性分析和多属性分析并没有显著区别。

本文根据上述的数据分析结果, 将中心频率作为自变量, 测点数据作为因变量, 建立多阶拟合模型。由于正常测点间并无显著差异, 因此可以任选一个测点作为模板, 分别建立模板数据及待诊断样本的多阶拟合模型, 计算两个系数向量之间的欧氏距离, 并通过统计方法设计适当的阈值, 当系数距离大于指定阈值时, 说明该测点所检测的设备很有可能出现异常, 反之设备正常运行。将测试数据集应用到上述方法, 故障诊断正确率高达 0.9, F 值为 0.89, 因此, 本文所提出的多阶拟合法具有实际意义。

本文创新性的提出多阶拟合法, 为螺杆泵的异常检测提供理论依据。但是仍存在一定的缺陷, 比如数据量有限, 在将来若能收集更多的数据就能够帮助我们得到统计上更显著的结论, 从而不断拓展实验思路。此外, 本文提出的多阶拟合法具有很强的可拓展性, 其研究思想和实验思路除了可以应用于螺杆泵的异常检测之外, 还能够应用于其他设备的故障诊断, 应用前景十分广阔。

基金项目

天津市智能制造专项资金项目 201810602, 201907206, 201907210; 天津市互联网先进制造专项资金项目 18ZXRHGX00110。

参考文献

- [1] 贾昭昭. 三螺杆泵性能评估及故障诊断研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2017.

- [2] Zaher, A., McArthur, S. and Infield, D. (2009) Online Wind Turbine Fault Detection through Automated SCADA Data Analysis. *Wind Energy*, **12**, 574-593. <https://doi.org/10.1002/we.319>
- [3] Zhang, S., Ye, F., Wang, B. and Habetler, T.G. (2019) Semi-Supervised Learning of Bearing Anomaly Detection via Deep Variational Autoencoders. arXiv:1912.01096 [cs. LG].
- [4] Guggilam, S., Zaidi, S.M.A., Chandola, V. and Patra, A.K. (2019) Integrated Clustering and Anomaly Detection (INCAD) for Streaming Data (Revised). In: *Lecture Notes in Computer Science*, Volume 11539, Springer, Berlin. https://doi.org/10.1007/978-3-030-22747-0_4
- [5] Han, J., *et al.* (2006) *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco, CA, 157-164.
- [6] Syakur, M.A., Syakur, M.A., Khotimah, B.K., *et al.* (2018) Integration K-Means Clustering Method and Elbow Method for Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, **336**, Article ID: 012017. <https://doi.org/10.1088/1757-899X/336/1/012017>
- [7] 张良均, 等. *Python 数据分析与挖掘实战*[M]. 北京: 机械工业出版社, 2016: 113-114
- [8] 茆诗松. *概率论与数理统计教程*[M]. 北京: 高等教育出版社, 2004.
- [9] 何春雄, 龙卫江, 朱锋峰. *概率论与数理统计*[M]. 北京: 高等教育出版社, 2012: 7.
- [10] 李萍, 王茂才, 林琳, 等. 最小二乘多项式拟合算法在管理高消耗医用低值耗材中的应用[J]. *中国卫生经济*, 2019, 38(11): 72-75.