



文章编号:1001-9081(2018)07-2136-05

DOI:10.11772/j.issn.1001-9081.2018010035

基于 Gibbs 抽样的轴承故障诊断方法

王 岩*, 罗 倩, 邓 辉

(北京信息科技大学 信息与通信工程学院, 北京 100101)

(* 通信作者电子邮箱 wyleo7@qq.com)

摘要:针对现有轴承故障诊断方法的不足,即诊断片面性问题,提出了一种基于 Gibbs 抽样的轴承故障诊断方法。首先对轴承振动信号进行局部特征尺度分解(LCD)得到内禀尺度分量(ISC);然后对轴承振动信号和 ISC 分别提取时域特征,按照特征敏感度高低对时域特征排名,选择排名靠前的特征组成特征集;其次使用特征集训练产生基于 Gibbs 抽样的多维高斯分布模型;最后通过后验分析得到概率,实现轴承故障诊断。实验结果表明诊断正确率达到 100%,与基于 SVM 的轴承诊断方法相比,在特征数为 43 个时诊断正确率提升了 11.1 个百分点。所提方法能够有效地对滚动轴承故障状态进行诊断,对高维复杂的轴承故障数据也有很好的诊断效果。

关键词:滚动轴承;故障诊断;局部特征尺度分解;多维高斯分布;吉布斯抽样

中图分类号: TP277 **文献标志码:**A

Bearing fault diagnosis method based on Gibbs sampling

WANG Yan*, LUO Qian, DENG Hui

(College of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing 100101, China)

Abstract: To suppress judgment one-sidedness in the existing bearing fault diagnosis method, a bearing fault diagnosis method based on Gibbs sampling was proposed. Firstly, the bearing vibration signal was decomposed by Local Characteristic Scale Decomposition (LCD) to obtain Intrinsic Scale Components (ISC). Secondly, the time domain features were extracted from the bearing vibration signal and ISC, and the time domain features were ranked according to feature sensitivity level. The top ranked features were selected to make up feature sets. Thirdly, feature set training was used to generate a multi-dimensional Gaussian distribution model based on Gibbs sampling. Finally, posterior analysis was used to obtain the probability to realize bearing fault diagnosis. The experimental results show that the diagnostic accuracy of the proposed method reaches 100%; compared with the bearing diagnosis method based on SVM (Support Vector Machine), the diagnostic accuracy is improved by 11.1 percentage points when the number of features is 43. The proposed method can effectively diagnose rolling bearing faults, and it also has good diagnostic effect on high-dimensional and complex bearing fault data.

Key words: rolling bearing; fault diagnosis; Local Characteristic scale Decomposition (LCD); multi-dimensional Gaussian distribution; Gibbs sampling

0 引言

在旋转机械中,滚动轴承是易于损坏的零部件之一。据有关资料统计,旋转机械的故障中振动故障占 70%,而 30% 的振动故障是由滚动轴承故障引起的^[1],因此滚动轴承故障诊断的理论和应用研究一直是一个重点。

随着大数据和人工智能时代的到来,越来越多的机器学习算法被用于轴承故障诊断中的模式识别,如 Ali 等^[2]采用神经网络,焦卫东等^[3]使用支持向量机(Support Vector Machine, SVM)进行轴承故障诊断,但它们或多或少地存在一些缺陷,比如支持向量机中针对每一种轴承故障都设置一个判别函数进行诊断,造成输入空间中存在无法分类的区域,即诊断片面性的问题^[4]。为解决此问题,本文提出了一种新的基于 Gibbs 抽样的轴承故障诊断算法,该方法将概率统计学应用到轴承故障诊断中。Gibbs 抽样是马尔可夫链蒙特卡

罗(Markov Chain Monte Carlo, MCMC)理论的一个分支,用来获取一系列近似等于指定多维概率分布(比如 2 个或者多个随机变量的联合概率分布)观察样本的算法。使用 Gibbs 抽样建立每一种轴承状态的分布模型,针对一个待诊断的样本,分别计算出属于每一种轴承状态的概率,选取最大的判定为该类别,从而达到故障诊断的目的,解决了轴承故障诊断中存在的诊断片面性问题。

基于 Gibbs 抽样的滚动轴承故障诊断思路如下:首先对轴承故障振动信号进行局部特征尺度分解(Local Characteristic scale Decomposition, LCD)得到若干个内禀尺度分量(Intrinsic Scale Component, ISC),再对轴承故障振动信号和 ISC 分别提取时域特征,对所有时域特征使用特征的类间标准差和类内标准差的比值筛选出敏感特征并组成特征集。针对不同种类的轴承故障,使用基于 Gibbs 抽样算法训练特征集产生不同的多维高斯分布模型。对于待诊断的故障

收稿日期:2018-01-08;修回日期:2018-03-08;录用日期:2018-03-14。

基金项目:国家自然科学基金资助项目(61271198);北京市科技提升计划项目(5211624101)。

作者简介:王岩(1993—),男,山东枣庄人,硕士研究生,主要研究方向:信号处理、大数据处理; 罗倩(1965—),女,广东兴宁人,副教授,博士,主要研究方向:信号处理、大数据处理; 邓辉(1994—),男,黑龙江哈尔滨人,硕士研究生,主要研究方向:信号处理、大数据处理。



数据, 分别计算在每个模型中的概率密度, 通过后验分析得到概率, 选取概率最大的结果将待诊断数据诊断为该故障类别。

1 数据预处理和特征提取

1.1 数据 LCD 预处理

准确提取轴承振动信号特征是轴承故障诊断中重要的一步, 如果直接从这些非平稳或非线性振动信号中提取特征势必影响故障诊断的效果, 因此特征提取前必须进行信号预处理。本文采用局部特征尺度分解进行信号预处理, 它能够同时在时域和频域提供非平稳信号的局部化信息, 且避免了经验模态分解 (Empirical Mode Decomposition, EMD) 的频率混淆问题^[5] 和局部均值分解 (Local Mean Decomposition, LMD) 的信号突变问题^[6], 所以它对非平稳信号处理效果比较好。

LCD 是一种基于极值点的局部尺度参数自适应的非平稳信号分解方法。该方法以任意相邻的两个极值点为跨度, 通过分段的形式对信号进行线性变换来构造基线函数, 从而将原始信号自适应地分解为若干个 ISC 分量。其具体步骤为: 将原始信号通过两个循环分解为具有物理意义的内禀尺度分量 ISC 和一个剩余项。第一个循环是寻找基线信号, 并经循环迭代求出 ISC 分量的过程, ISC 分量应满足 ISC 分量条件; 第二个循环是将原始信号减掉已获得的 ISC 分量后的剩余信号作为原始信号再次经循环迭代得到 ISC 分量。重复上述步骤直至所得剩余项为一个单调函数为止^[7]。对于任一时间序列, 通过 LCD 分解可以得到若干个 ISC 和一个剩余项。LCD 分解本质是对非平稳信号进行平稳化处理, 将信号中不同尺度的波动或趋势逐级分解。因为信号的能量主要集中在前几个 ISC 分量中, 因此本文只取前 4 个 ISC 分量。

1.2 特征提取

在 LCD 预处理之后, 对原始振动信号和 4 个 ISC 分量分别进行时域特征提取。故障特征可以分为无量纲特征和有量纲特征两大类^[8-9]。无量纲指标因其不受工况影响的特点, 对剥落、压痕等故障有很好的诊断效果; 有量纲指标对磨损类故障诊断效果比较好。因此, 为了能够更加全面、准确地诊断故障, 本文对原始信号和 4 个 ISC 信号分别提取了 6 个无量纲特征(偏斜度、峭度、峰值、波形指标、脉冲、裕度)和 5 个有量纲特征(有效值、均值、标准差、方根幅值、最大值), 共得到 55 个特征。

1.3 特征筛选方法

如果使用上述 55 个特征作为一个特征集, 计算量过大且诊断效果不一定最佳, 所以需要从中挑选出敏感特征用于模型训练。本文使用了基于故障类间、类内标准差比值的特征筛选方法。

类内标准差计算如下:

$$S_o = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{1}{K-1} \sum_{k=1}^K (S_k^n - \bar{S}^n)^2} \quad (1)$$

其中: S^n 表示某一特征的特征点集合, \bar{S}^n 表示特征点的均值, K 表示特征点个数, N 表示轴承故障的状态个数, S_o 表示类内标准差。

类间标准差计算如下:

$$S_i = \sqrt{\frac{1}{N-1} \sum_{k=1}^K \left(\frac{1}{K} \sum_{n=1}^N S_k^n - \bar{Y} \right)^2} \quad (2)$$

其中: \bar{Y} 表示某一特征在每种轴承状态下的均值, S_i 表示类间标准差。

对于某一特征, 其类间标准差与类内标准差的比值, 称为该特征的敏感度指标。计算公式如下:

$$\chi = S_i / S_o \quad (3)$$

其中: χ 表示敏感度指标, 特征对不同状态的区别度与敏感度成正相关, 某一特征的 χ 值越大, 该特征对不同状态的区别度越高。

2 基于 Gibbs 抽样的轴承故障诊断

本文只涉及到轴承运行的 4 种状态: 内圈故障状态、外圈故障状态、滚动体故障状态和正常状态。基于 Gibbs 抽样的轴承故障诊断基本思路是: 从上述 55 个特征中选取排名靠前的若干敏感特征组成特征集, 针对轴承的每一个状态, 对这些特征建立多维概率分布, 使用 Gibbs 算法对多维概率分布进行参数估计, 最终得到 4 个多维概率分布模型。对于一个待诊断轴承数据, 分别计算在 4 种模型中的概率密度, 通过后验分析得到概率, 选取最大概率判定待诊断数据为该轴承状态类别。

2.1 特征值分布

建立特征的多维概率分布, 首先需要分析每个特征在不同状态下的一维分布, 然后构建出合适的多维概率分布。生成轴承不同状态下各特征的分布直方图, 通过直观观察发现绝大多数直方图服从高斯分布。进一步, 使用 Quantile-Quantile 正态检验方法^[10] 和 Michael 拟合优度检验方法^[11] 检验各特征分布, 发现除信号峭度特征在任何轴承状态下不满足高斯分布外, 其他特征都服从高斯分布, 这些特征都落在 90% 接受区间内, 可以认为这些特征都服从高斯分布, 因此, 针对轴承的每一个状态, 可以建立特征集的多维高斯分布。

2.2 多维高斯分布参数估计

使用若干个排名靠前的特征组成特征集构建多维高斯分布, 对于轴承故障的第 j 个状态的多维高斯分布模型 V_j 的参数有均值 μ_j 、协方差矩阵 Σ_j , 对这两个参数进行参数估计。高斯分布均值 μ_j 的先验概率服从高斯分布, 协方差矩阵 Σ_j 的先验概率服从逆 Wishart 分布, 通过对参数先验分布进行 Gibbs 抽样, 可以得到这两个参数具体的值^[12]。故首先对每个多维高斯分布分别建立参数的先验分布。

使用 Gibbs 算法的流程为针对每一个轴承状态的多维高斯分布, 分别从参数的先验分布中随机抽样得到 $\mu_j^{(n)}, \Sigma_j^{(n)}$, 右上标 n 表示第 n 次抽样。通过多次抽样后, 比如 N 次抽样, 为保证结果收敛 N 至少取 1000, 参数 μ_j 和 Σ_j 都可以得到 N 次抽样结果。通常抽样结果中初始部分的数据点是不满足收敛的, 故舍去前 $N/2$ 次抽样结果, 即舍去收敛之前被称为 burn-in period 阶段的抽样数据^[13]。保留剩下后 $N/2$ 次抽样结果, 然后对每个参数求平均值, 最终可以得到每个多维高斯分布参数近似的期望和协方差。假设选择 D 个特征组成训练特征集, 第 j 个状态多维高斯分布参数估计具体步骤如下。

1) 从逆 Wishart 先验分布中随机抽样协方差矩阵 Σ_j , 下标 j 表示轴承第 j 个状态:

$$\Sigma_j^{(n)} | Y_j \sim W^{-1}(n_j + f_j, \psi_j + Q_j + \frac{n_j \tau_j}{n_j + \tau_j} (\bar{y}_j - \xi_j)(\bar{y}_j - \xi_j)^T) \quad (4)$$



其中: f_j 为逆 Wishart 分布的自由度初始参数, ψ_j 为逆 Wishart 分布的初始尺度矩阵 (Scale matrix) 是一个 $D \times D$ 的矩阵, \mathbf{Y}_j 为训练样本提取并筛选出的特征集, n_j 表示每个特征中包含特征点个数, ξ_j 为特征集中心点的初始坐标, $\bar{\mathbf{y}}_j$ 表示特征集中特征的平均值, τ_j 表示精度参数。 \mathbf{Q}_j 计算公式为:

$$\mathbf{Q}_j = \sum_{i \in n_j} (\mathbf{y}_i - \bar{\mathbf{y}}_j)(\mathbf{y}_i - \bar{\mathbf{y}}_j)^T \quad (5)$$

2) 从高斯先验分布中抽样协方差矩阵 $\boldsymbol{\mu}_j$, 如式(6):

$$\boldsymbol{\mu}_j^{(n)} + \mathbf{Y}_j, \boldsymbol{\Sigma}_j^{(n)} \sim N(\bar{\xi}_j, \boldsymbol{\Sigma}_j^{(n)} / (\tau_j + m_j)) \quad (6)$$

其中 $\bar{\xi}_j$ 的更新公式为:

$$\bar{\xi}_j = (\tau_j \xi_j + n_j \bar{\mathbf{y}}_j) / (n_j + \tau_j) \quad (7)$$

第 j 个轴承状态的多维高斯分布参数估计的 Gibbs 算法通过将上述两个步骤运算 N 次, 可以分别得到每个参数的 N 个抽样结果, 最终 $\boldsymbol{\mu}_j$ 和 $\boldsymbol{\Sigma}_j$ 参数值由式(8)~(9) 得到, 其中舍弃了前 $N/2$ 的抽样结果^[13]。

$$\boldsymbol{\mu}_j = \frac{1}{N/2} \left(\sum_{n=N/2}^N \boldsymbol{\mu}_j^{(n)} \right) \quad (8)$$

$$\boldsymbol{\Sigma}_j = \frac{1}{N/2} \left(\sum_{n=N/2}^N \boldsymbol{\Sigma}_j^{(n)} \right) \quad (9)$$

轴承故障第 j 个状态由均值 $\boldsymbol{\mu}_j$ 和协方差矩阵 $\boldsymbol{\Sigma}_j$ 构成的多维高斯分布模型为 V_j :

$$u_j = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}_j|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{X} - \boldsymbol{\mu}_j) \right\} \quad (10)$$

其中: \mathbf{X} 是待诊断数据根据训练特征集提取的时域特征; D 是特征集中包含特征的个数, 即多维高斯分布的维数; u_j 表示 \mathbf{X} 在第 j 个轴承状态模型中计算得到概率密度。

得到概率密度后, 需要进行后验概率分析。在轴承故障诊断之前, 假设该轴承被诊断为 4 种状态的先验概率 Z_j 是相等的, 即 $Z_j = 0.25$ 。由贝叶斯定理^[14] 可得 \mathbf{X} 被诊断为第 j 个状态的后验概率 p_j 为:

$$p_j = Z_j \times u_j / \left(\sum_{j=1}^4 (Z_j \times u_j) \right) \quad (11)$$

2.3 轴承故障诊断

使用轴承不同状态的振动信号样本数据经由 LCD 并提取时域特征后, 选择 D 个敏感特征组成特征集, 训练产生 4 个多维高斯分布模型。对于待诊断的数据, 先根据训练特征集提取的时域特征分别计算式(10) 得到概率密度 u_1, u_2, u_3 和 u_4 , 再使用式(11) 分别计算概率 p_1, p_2, p_3 和 p_4 , 选取最大概率诊断为该待诊断数据的类别。

3 仿真及实测数据诊断结果分析

使用了凯斯西储大学的滚动轴承振动数据进行仿真(数据来源: <http://csegroups.case.edu/Bearingdatacenter/pages/download-datafile>)。振动信号由驱动端加速度传感器采集, 轴承型号为 SKF6205, 电机转速为 1772 r/s, 使用电火花加工技术在驱动端轴承上布置了单点故障, 内圈、滚动体及外圈故障直径都为 0.007 inch(1 inch = 2.54 cm), 外圈故障位于 6 点钟位置, 采样频率为 12 kHz, 实际故障频率: 内圈为 159.93 Hz, 滚动体为 139.20 Hz, 外圈为 105.88 Hz。训练模型时, 将轴承振动信号若干个采样点作为一组, 从凯斯西储大学原始数据

中提取一共 240 组组成训练样本, 其中轴承 4 种状态各占 60 组。测试模型时, 从原始数据中另外提取一共 240 组组成测试数据, 其中轴承 4 种状态也各占 60 组。将训练数据的时域特征按照敏感度从高到低排列, 其中排名前 10 的特征如表 1 所示, 逐个将特征加入特征集进行模型训练, 可以得到 55 组不同特征集的模型。测试数据按照训练特征集提取特征, 分别计算其在 55 组不同模型中的诊断正确率, 可以得到特征集取不同数量的特征时该算法的诊断正确率。

表 1 特征敏感度排名

Tab. 1 Feature sensitivity ranking

排名	特征名称	排名	特征名称
1	第二 ISC 最大值	6	原始信号有效值
2	原始信号最大值	7	原始信号标准差
3	第二 ISC 标准差	8	第一 ISC 有效值
4	第二 ISC 有效值	9	第一 ISC 标准差
5	第一 ISC 最大值	10	第一 ISC 方根幅值

从图 1 中可以看出, 当特征集的特征个数增加到 2 时, 诊断正确率就达到 100%。图 1 中的诊断正确率随着特征数的增加呈下降趋势, 因为特征按照本文提到的特征筛选排名方法, 逐个顺序加入特征集并计算正确率。按此排名顺序, 特征对于区分不同轴承故障状态的敏感度随着特征数增加而下降, 第 20 个特征之后加入的特征敏感度较差, 会对整体诊断正确率产生影响, 因此图 1 中诊断正确率随着特征数的增加而下降, 也验证了特征筛选排名方法的有效性。

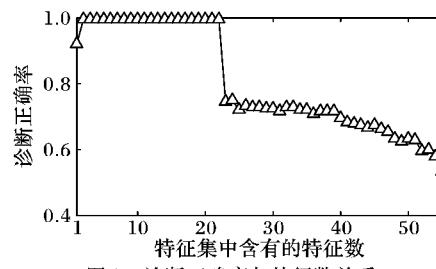


图 1 诊断正确率与特征数关系

Fig. 1 Relationship between accuracy of diagnosis and features

将 240 组测试集数据分别代入使用 4 个特征作为特征集的训练模型中, 按照诊断结果的标签对 4 维数据使用了多维标度法 (Multi-Dimensional Scaling, MDS)^[15] 映射在二维空间中, 如图 2 所示。

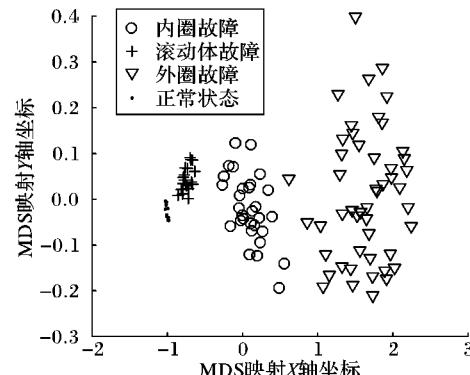


图 2 高维测试集数据的二维映射 (240 组测试集数据)

Fig. 2 Two dimensional mapping of high dimensional test set data (240 sets of testing set data)



从图 2 中可以看出,4 种状态数据特征分布较为鲜明,说明本文提出的算法可以很好地判别 4 种不同的轴承状态。

对前 4 个特征组成的特征集进一步分析,使用上述 240 组训练集产生轴承 4 种不同状态的模型: 内圈故障 V_1 、外圈故障 V_2 、滚动体故障 V_3 和正常状态 V_4 。每个模型的参数 μ_j 和 Σ_j 计算结果列于表 2, 在图 3 中使用多维标度法将不同模型的参数 μ_j 映射在二维空间中。

表 2 多维高斯分布模型参数

Tab. 2 Parameters of multidimensional Gaussian distribution model

模型	Σ	μ
V_1	$\begin{bmatrix} 0.0276 & 0.0274 & 0.0274 & 0.0274 \\ 0.0274 & 0.0275 & 0.0274 & 0.0274 \\ 0.0274 & 0.0274 & 0.0275 & 0.0274 \\ 0.0274 & 0.0274 & 0.0274 & 0.0275 \end{bmatrix}$	$\begin{bmatrix} 0.3145 \\ 1.1246 \\ 0.1197 \\ 0.1195 \end{bmatrix}$
V_2	$\begin{bmatrix} 0.0034 & 0.0034 & 0.0034 & 0.0034 \\ 0.0034 & 0.0034 & 0.0034 & 0.0034 \\ 0.0034 & 0.0034 & 0.0034 & 0.0034 \\ 0.0034 & 0.0034 & 0.0034 & 0.0034 \end{bmatrix}$	$\begin{bmatrix} 0.0889 \\ 0.3861 \\ 0.0427 \\ 0.0426 \end{bmatrix}$
V_3	$\begin{bmatrix} 0.1056 & 0.1052 & 0.1052 & 0.1052 \\ 0.1052 & 0.1057 & 0.1053 & 0.1053 \\ 0.1052 & 0.1053 & 0.1057 & 0.1053 \\ 0.1052 & 0.1053 & 0.1053 & 0.1057 \end{bmatrix}$	$\begin{bmatrix} 0.6772 \\ 2.4354 \\ 0.2460 \\ 0.2454 \end{bmatrix}$
V_4	$\begin{bmatrix} 1.0E-3 * \\ 0.7213 & 0.7184 & 0.7182 & 0.7185 \\ 0.7184 & 0.7214 & 0.7183 & 0.7185 \\ 0.7182 & 0.7183 & 0.7212 & 0.7184 \\ 0.7185 & 0.7185 & 0.7184 & 0.7217 \end{bmatrix}$	$\begin{bmatrix} 0.0604 \\ 0.1728 \\ 0.0295 \\ 0.0295 \end{bmatrix}$

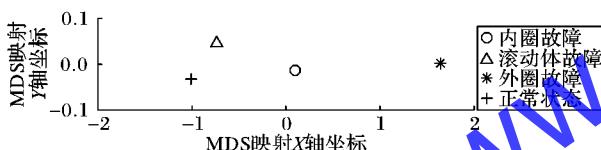


图 3 高维高斯分布均值的二维映射

Fig. 3 Two dimensional mapping of high-dimensional Gaussian distribution mean

为了验证本文提出的算法优于基于向量机的轴承故障诊断算法, 使用凯斯西储大学轴承数据中采样频率为 48 kHz 的数据, 振动信号由风扇端加速度传感器采集, 轴承型号为 SKF6205, 电机转速为 1797 r/s, 使用电火花加工技术在驱动端轴承上布置了单点故障, 内圈、滚动体及外圈故障直径都为 0.021 英寸, 外圈故障位于 3 点钟位置, 实际故障频率: 内圈为 162.20 Hz, 滚动体为 141.09 Hz, 外圈为 107.30 Hz。使用本文中的特征提取、筛选及排名的方法, 在轴承故障诊断中分别使用了基于 Gibbs 抽样的诊断方法和基于支持向量机的诊断方法。原始数据中缺少采样频率为 48 kHz 的轴承正常状态数据, 因此只建立了有关内圈故障、外圈故障及滚动体故障的模型, 从凯斯西储大学原始数据中提取一共 180 组数据组成训练样本, 其中轴承 3 种状态各占 60 组。测试模型时, 从原始数据中另外提取一共 180 组组成测试数据, 轴承 3 种状态也各占 60 组。特征按照本文中的方法进行排名并依次顺序加入特征集, 计算每个特征集有关模型的诊断正确率。

如图 4 所示: 在特征数比较少时, 两种诊断方法都能达到最高 100% 的诊断正确率; 随着特征数的增加, 特征对轴承不同状态的敏感度降低, 二者的诊断正确率都下降, 但基于

Gibbs 抽样诊断算法下降的幅度明显比支持向量机小; 特别在特征数达到 43 时, 基于 Gibbs 抽样方法诊断正确率为 82.8%, 而基于支持向量机方法诊断正确率为 71.7%。与基于支持向量机诊断算法相比, 基于 Gibbs 抽样诊断算法正确率提升了 11.1 个百分点; 当特征数大于 43 后, 两种算法的诊断正确率差值进一步增大。说明 Gibbs 抽样诊断算法对于敏感度低的特征诊断效果好于支持向量机, 特别对于高维复杂数据的离群点及支持向量机输入空间中存在无法分类的区域^[4], Gibbs 抽样诊断算法通过概率计算依然可以实现对轴承故障的诊断。

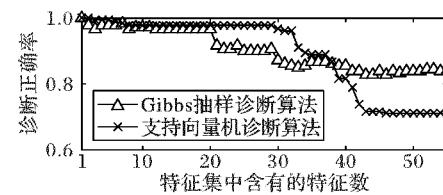


图 4 两种算法的诊断正确率比较

Fig. 4 Diagnostic accuracy comparison of two algorithms

凯斯西储大学的数据是实验室条件下得到的数据, 与真实的数据还有一定差别。使用该方法对中国铁路某局实际的滚动轴承数据进行了诊断。数据来自中国铁路某局甲 2014 年某月。该数据只有轴承故障和正常两个状态, 并没有对故障进行细分, 数据是经过共振解调^[15]预处理过的。图 5 是其振动信号 200 个抽样点的时间序列波形。

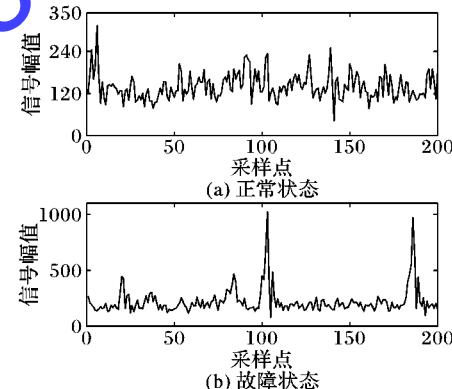


图 5 轴承振动信号时间序列

Fig. 5 Bearing vibration signal time series

使用 60 组故障数据和 60 组正常数据作为训练样本, 另外 60 组故障数据和 60 组正常数据作为测试数据。图 6 是该局使用本文提出的诊断方法在不同数量特征集下的诊断正确率, 由图可知取 14 个特征组成特征集时诊断正确率最大, 为 95.8%。

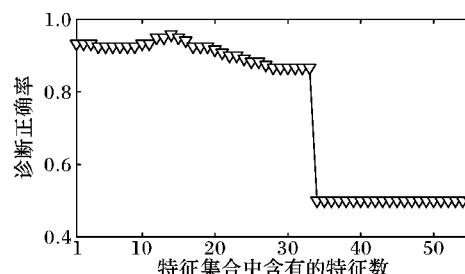


图 6 某局甲的故障诊断正确率

Fig. 6 Fault diagnosis accuracy of a railway bureau



使用14个特征组成特征集对测试集数据进行诊断,诊断结果的标签使用了多维标度法(Multi-Dimensional Scaling, MDS)^[15]显示二维空间中,如图7所示。从实际数据仿真结果可以看出,本文方法依然有较高的故障诊断正确率。

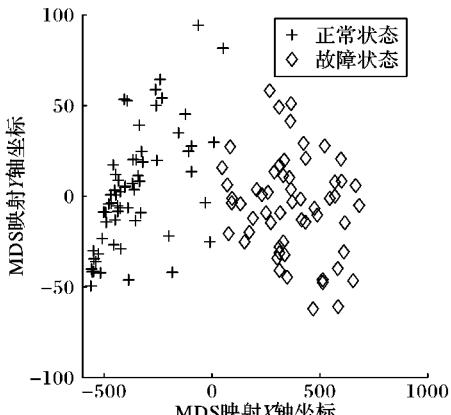


图7 高维测试集数据的二维映射(14个特征组成特征集)

Fig. 7 Two dimensional mapping of high dimensional test set data (feature set of 14 features)

4 结语

本文提出了一种基于Gibbs抽样的轴承故障诊断算法,通过理论分析及数据仿真结果分析表明:

1) 本文算法将概率统计学运用到轴承故障诊断中,采用概率模型代替支持向量机中的判别函数进行故障诊断,参考每种模型概率密度的计算结果,通过后验分析得到概率,以判别概率最大为准则,对于支持向量机输入空间中无法分类的区域^[4]也能有效判别,避免了判别偶然性和片面性的出现,使得识别结果更加客观。与基于SVM的轴承诊断方法相比,在特征数为43时诊断正确率提升了11.1个百分点。

2) 提出了将Gibbs抽样算法应用于滚动轴承故障诊断,可以实现从信号处理到分类的智能故障诊断。

3) 本文分别使用凯斯西储大学实验室条件下的数据和中国铁路机车轴承的实际采集数据进行了算法验证,结果表明本文方法能够有效地诊断轴承故障状态。

参考文献 (References)

- [1] 王小龙.滚动轴承故障诊断研究[J].成都大学学报(自然科学版),2016,35(2):178-182.(WANG X L. Research of fault diagnosis in rolling bearing [J]. Journal of Chengdu University (Natural Science), 2016, 35(2): 178 - 182.)
- [2] ALI J B, FNAIECH N, SAIDI L, et al. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals [J]. Applied Acoustics, 2015, 89(3): 16 - 27.
- [3] 焦卫东,林树森.整体改进的基于支持向量机的故障诊断方法[J].仪器仪表学报,2015,36(8):1861-1870.(JIAO W D, LIN S S. Overall-improved fault diagnosis approach based on support vector machine [J]. Chinese Journal of Scientific Instrument, 2015, 36 (8): 1861 - 1870.)
- [4] BISHOP C M. Pattern Recognition and Machine Learning (Information Science and Statistics) [M]. Berlin: Springer, 2006: 182 - 184.
- [5] 江莉,李林,董惠.基于改进EMD方法的多分量信号分析[J].振动与冲击,2009,28(4):51-53.(JIANG L, LI L, DONG H. An improved empirical mode decomposition method for multicomponent signal representation [J]. Journal of Vibration and Shock, 2009, 28 (4): 51 - 53.)
- [6] LEI Y, HE Z, ZI Y. Application of the EEMD method to rotor fault diagnosis of rotating machinery [J]. Mechanical Systems & Signal Processing, 2009, 23(4): 1327 - 1338.
- [7] 吴占涛,程军圣,李宝庆,等.广义局部特征尺度分解方法及其应用[J].振动工程学报,2016,29(2):331-339.(WU Z T, CHENG J S, LI B Q, et al. The method of generalized local characteristic-scale decomposition and its application [J]. Journal of Vibration Engineering, 2016, 29(2): 331 - 339.)
- [8] 陈俊君,徐冰.基于无量纲指标与小波变换的轴承故障诊断研究[J].煤矿机械,2015,36(5):314-316.(CHEN J J, XU B. Study of fault diagnosis for bearing based on non-dimensional parameter and wavelet transform [J]. Coal Mine Machinery, 2015, 36 (5): 314 - 316.)
- [9] HE J, XUAN J F, REN Z L, et al. Misleading classification [J]. Science China Information Sciences, 2014, 57(5): 1 - 17.
- [10] EASTON G, MCCULLAGH P. A multivariate generalization of quantile-quantile plots [J]. Publications of the American Statistical Association, 1990, 85(410): 376 - 386.
- [11] MICHAEL J R. The stabilized probability plot [J]. Biometrika, 1983, 70(1): 11 - 17.
- [12] LAN F, PAN Y, CHENG K T T. An efficient network-on-chip yield estimation approach based on Gibbs sampling [J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2016, 35(3): 447 - 457.
- [13] 刘景森,金勇.一种基于吉布斯抽样的MUSIC多维参数联合估计算法[J].传感器与微系统,2008,27(6):62-65.(LIU J S, JIN Y. New Gibbs sampling multi-parameters joint estimator based on MUSIC method [J]. Transducer and Microsystem Technologies, 2008, 27(6): 62 - 65.)
- [14] FRANZÉN J. Bayesian cluster analysis some extensions to non-standard situations [J]. Biometrika, 2008, 65(1): 31 - 38.
- [15] 屈天国,蔡自兴.基于分而治之的多维标度算法[J].模式识别与人工智能,2014,27(11):961-969.(QU T G, CAI Z X. A divide-and-conquer based multidimensional scaling algorithm [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(11): 961 - 969.)

This work is partially supported by the National Natural Science Foundation of China (61271198), the Beijing Municipal Science and Technology Project (5211624101).

WANG Yan, born in 1993, M. S. candidate. His research interests include signal processing, big data processing.

LUO Qian, born in 1965, Ph. D., associate professor. Her research interests include signal processing, big data processing.

DENG Hui, born in 1994, M. S. candidate. His research interests include signal processing, big data processing.