

DOI: 10.13382/j.jemi.B2003275

基于隐 Markov 模型的齿轮箱故障识别方法研究*

杨秀芳 陈 卓 王 驰
(西安理工大学 机械与精密仪器工程学院 西安 710048)

摘 要:针对神经网络的识别一直停留在静态模式识别上的不足,采用一种近年来发展较快的动态模式识别技术—隐马尔科夫模型分析齿轮箱振动信号。首先提取齿轮箱振动信号在时域、频域和时频域的统计特征,组成 34 维全特征矢量,训练了一组全特征-隐马尔科夫模型库;再通过主分量分析技术对全特征矢量进行降维处理,取其吸收信息量 98% 以上的前 7 个主分量组成主分量特征矢量,训练了另外一组主分量-隐马尔科夫模型库。分别用两组独立的模型库进行齿轮箱故障识别。结果表明,全特征-隐马尔科夫模型库对齿轮正常、齿轮断齿的识别准确率为 97.9%,对齿轮点蚀的识别准确率为 100%,程序运行耗时 22.328 s,主分量-隐马尔科夫模型库对齿轮点蚀的和齿轮断齿识别准确率均达到 100%,程序运行耗时 4.879 s。振动信号特征的降维处理没有降低故障的识别率,反而提升了故障识别准确率,同时大大提升了程序运行速度,这对机械系统故障诊断具有重要意义。

关键词: 故障识别; 齿轮箱; 隐马尔科夫模型; 全特征矢量; 主分量特征矢量

中图分类号: TP277;TN98;TH165+.3 **文献标识码:** A **国家标准学科分类代码:** 520.20

Research on gearbox fault identification method based on hidden Markov model

Yang Xiufang Chen Zhuo Wang Chi
(Faculty of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: Aiming at the shortcomings of neural network recognition in static pattern recognition, a dynamic pattern recognition technology developed in recent years—hidden Markov model is used to analyze gearbox vibration signals. First, the statistical characteristics of the gearbox vibration signals in the time domain, frequency domain and time-frequency domain are extracted to form a 34-dimensional full feature vector. Trained a set of full feature-hidden Markov model libraries; then, through the principal component analysis technology, the full feature vector is reduced in dimension, and the first 7 principal components whose absorption information is more than 98% constitute the principal component feature vector. Another set of principal component-hidden Markov model library was trained. Two sets of independent model libraries are used for gearbox fault identification. The results show that the full feature-hidden Markov model library has 97.9% accuracy for the identification of normal gears and gear broken tooth and 100% for gear pitting. The program takes 22.328 s. The recognition accuracy of component-hidden Markov model library for gear pitting and gear tooth failure is 100%. The program takes 4.879 s. Therefore, the dimensionality reduction processing of the vibration signal feature does not reduce accuracy of fault identification, but improves the accuracy of fault recognition, and greatly increases the speed of the program. This is of great significance for fault diagnosis of mechanical systems.

Keywords: fault identification; gearbox; hidden Markov model; full feature vector; principal component feature vector

0 引言

目前对齿轮箱故障诊断的研究主要集中在齿轮箱的状态检测仪器及分析系统的开发、信号处理与分析、故障机理研究与典型故障特征提取、诊断及人工智能应用等方面。隐马尔科夫模型(hidden Markov model, HMM)是引入故障诊断技术的新方法,适合对信息量大、非平稳、特征重复性不佳的诊断信号进行处理,在模式识别特别是在语音识别领域取得了很大成功^[1-3]。

20 世纪 70 年代, HMM 基本理论, 由 Baum 等^[4-5]首次建立, 20 世纪 80 年代, Rabiner^[6]将 HMM 基本理论引入到语音识别领域, 取得了令人满意的结果, 使得 HMM 技术成为语音识别领域最有效的技术之一。由于 HMM 在模式识别方面的良好表现, 以及可靠的计算性能和良好的可解释性, 近年来迅速扩展到下肢假肢运动意图识别^[7]、人脸检测^[8]、手语识别^[9]、车牌识别^[10]和视频监控^[11]。在机械制造、工业生产、航空航天等领域, 可用于故障诊断和预测, Heck 等^[12]首次将隐 Markov 模型引入到机械系统的状态监测中, 采用连续马尔科夫模型建立了工具的磨损模型, Bunks 等^[13]通过对基于隐 Markov 模型的语音识别方法的研究, 给出了预测剩余寿命的理论和方法, 并通过直升机齿轮箱的故障诊断实验进行了验证, 该研究意义深远, 具有重大的推动作用。在国内, 浙江大学的吴昭同团队最早研究将隐 Markov 模型用于旋转机械的故障诊断中, 童进等^[14]采用 HMM-AR 对大型旋转机械升速过程进行故障诊断, 取得了较好的诊断效果; 冯长建等^[15]针对旋转机械启动过程中的特征提取技术及标量化方法、多观测样本下的连续隐 Markov 模型训练算法的改进等做了一系列研究; 张西宁等^[16]提出自适应基因粒子群算法优化隐马尔科夫模型, 提高了滚动轴承故障诊断准确性。

鉴于 HMM 技术在齿轮箱故障诊断中应用比较少, 本文采用主分量分析法(principal component analysis, PCA)对齿轮箱振动信号的特征进行降维处理, 用降维前和降维后的特征分别训练 2 组 HMM 库, 并应用于齿轮箱故障诊断中, 验证了特征向量对 HMM 训练和对故障诊断的影响。

1 HMM 基本理论和算法

1.1 HMM 基本理论

HMM 是在 Markov 链的基础上发展起来的。它是一个双重随机过程, 包括 Markov 链和状态与观察变量之间的随机过程两部分组成, 其中马尔科夫链用于描述状态之间的转移, 一般随机过程描述状态和观测变量之间的统计关

系。基本的 HMM 组成如图 1 所示, 其中的 N 表示状态数, T 表示观察序列长度。 π 是状态初始概率分布矢量, A 是一个 $N \times N$ 的状态转移概率矩阵, B 是一个 $N \times M$ 的观察值概率矩阵, 即观察值对应某种状态的概率。

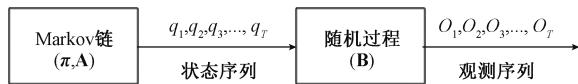


图 1 HMM 基本组成

Fig. 1 Basic composition of HMM

通常一个 HMM 简记为 $\lambda = (\pi, A, B)$

在实际应用中, 与 HMM 相关的基本问题有 3 种。

1) 评估(也称为分类), 在给定的观察序列 $O = O_1, O_2, \dots, O_T$ 和参数组 $\lambda = (\pi, A, B)$ 的 HMM 中, 计算观察序列出现的概率。利用前向-后向算法解决这一问题。

2) 解码(又称为识别), 在给定的观察序列 $O = O_1, O_2, \dots, O_T$ 和参数组 $\lambda = (\pi, A, B)$ 的 HMM 中, 隐藏状态序列 $S = S_1, S_2, \dots, S_N$ 生成所给出的观察序列的概率是多少。使用 Viterbi 算法。

3) 学习(也称为训练), 如何调整模型中的参数 $\lambda = (\pi, A, B)$, 使出现给定观察序列的概率最大化。为了解决这个问题, 一种解决方案, 通过使用 Baum-Welch 算法。

1.2 HMM 算法

1) 前向-后向算法

这个算法解决 HMM 的评估问题。对于观察变量 $O = \{O_1, O_2, \dots, O_T\}$, 在给定模型 $\lambda = (\pi, A, B)$ 下, 概率 $P(O|\lambda)$ 的直接求取方法如下^[17]:

对于一个固定的状态序列 $S = q_1, q_2, \dots, q_T$, 对应观察变量的概率:

$$P(O|S, \lambda) = \prod_{t=1}^T P(O_t | q_t, \lambda) = b_{q_t}(O_1) b_{q_2}(O_2) \cdots b_{q_T}(O_T) \quad (1)$$

其中:

$$b_{q_t}(O_t) = b_{jk} | q_t = \theta_j, o_t = v_k, 1 \leq t \leq T \quad (2)$$

对给定的模型 λ , 产生 S 的概率为:

$$P(S|\lambda) = \pi_{q_1} \alpha_{q_1 q_2} \cdots \alpha_{q_{T-1} q_T} \quad (3)$$

最终概率为:

$$P(O|\lambda) = \sum_S P(O|S, \lambda) P(S|\lambda) = \sum \pi_{q_1} b_{q_1}(O_1) \alpha_{q_1 q_2} b_{q_2}(O_2) \cdots \alpha_{q_{T-1} q_T} b_{q_T}(O_T) \quad (4)$$

由于式(4)的计算工作量庞大, Baum 等^[4-5]提出的前向-后向算法解决了计算量的问题。

(1) 前向算法

在 t 时刻, 观察变量为 O_t , 状态 θ_i 的概率为:

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, q_t = \theta_i | \lambda), 1 \leq t \leq T \quad (5)$$

式中: $\alpha_t(i)$ 为前向概率变量。

在 $t+1$ 时刻由状态 i 转换为状态 j , 有观察变量 O_{t+1}

的概率:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}),$$

$$1 \leq t \leq T-1; 1 \leq j \leq N \quad (6)$$

最终概率为:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (7)$$

(2) 后向算法

与前向算法相类似,在 $t+1$ 时刻,观察变量为 O_{t+1} , 状态 θ_i 的概率为:

$$\beta_{t+1}(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | q_{t+1} = \theta_i, \lambda),$$

$$1 \leq t \leq T-1 \quad (8)$$

式中: $\beta_{t+1}(i)$ 为后向概率变量。

由 $t+1$ 时刻到 t 时刻,状态由 i 转换为状态 j ,观察变量由 O_{t+1} 转换的 O_t 的概率:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j),$$

$$t = T-1, T-2, \dots, 1; 1 \leq i \leq N \quad (9)$$

最终概率为:

$$P(O|\lambda) = \sum_{i=1}^N \beta_1(i) \quad (10)$$

前向算法和后向算法在 t 时刻与 $t+1$ 时刻的迭代关系如图 2 所示。

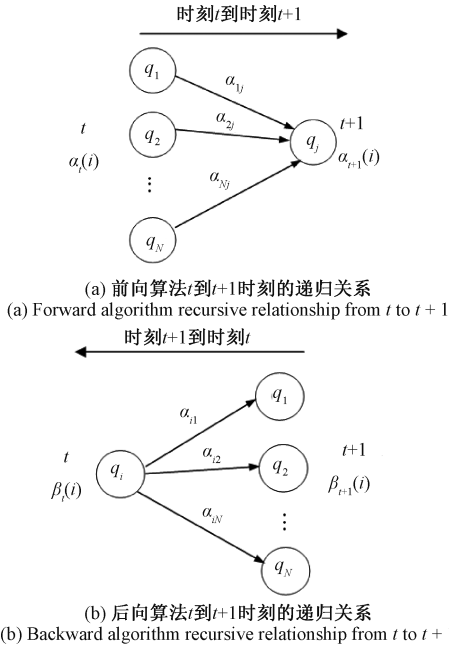


图 2 前向、后向算法在 t 与 $t+1$ 时刻的递归关系

Fig. 2 Recursive relation of forward and backward algorithm at time t and $t+1$

2) Viterbi 算法

这个算法解决 HMM 的解码 (或称为识别) 问题。对于观察变量 $O = \{O_1, O_2, \dots, O_T\}$, 在给定模型 $\lambda = (\pi, A, B)$ 下, 选择一个能够最好解释“观测变量”的状态序列

$Q = \{q_1, q_2, \dots, q_T\}$, 即概率 $P(Q|O, \lambda)$ 最大的状态序列。

定义 $\delta_t(i)$ 为 t 时刻沿路径 $q_1, q_2, \dots, q_T (q_t = \theta_i)$ 产生观测值序列 $O_1, O_2, \dots, O_T$ 的最大概率, 则:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_t} P(q_1, q_2, \dots, q_t, q_t = \theta_i, o_1, o_2, \dots, o_t | \lambda),$$

$$1 \leq t \leq T \quad (11)$$

对应的最佳状态:

$$q_t^* = \text{argmax}_{1 \leq i \leq N} [\delta_t(i)] \quad (12)$$

其中, 符号 argmax 定义为当 $\delta_t(i)$ 最大时, 对应自变量的值, 即对应的状态。

3) Baum-Welch 算法

这个算法解决 HMM 的训练问题。已知观察变量 $O = \{O_1, O_2, \dots, O_T\}$ 和初始条件, 调整模型 $\lambda = (\pi, A, B)$, 使得 $P(O|\lambda)$ 达到最大。Baum-Welch 算法通过递归运算使 $P(O|\lambda)$ 达到局部最大, 得到 HMM 模型参数。

定义 $\xi_t(i, j)$ 为给定训练序列 O 及模型 λ 时, 马尔科夫模型链在 t 时刻和 $t+1$ 时刻处于 θ_i 状态和 θ_j 状态的概率为:

$$\xi_t(i, j) = P(O, q_t = \theta_i, q_{t+1} = \theta_j | \lambda) \quad (13)$$

根据前向变量和后向变量的定义可导出:

$$\xi_t(i, j) = [\alpha_t(i) a_{ij} b_j(o_{t+1})] / P(O | \lambda) \quad (14)$$

那么, t 时刻马尔科夫链处于状态 θ_i 时的概率为:

$$\xi_t(i) = P(O, q_t = \theta_i | \lambda) =$$

$$\sum_{j=1}^N \xi_t(i, j) = \alpha_t(i) \beta_t(i) P(O | \lambda) \quad (15)$$

因此, $\sum_{i=1}^{T-1} \xi_t(i)$ 表示的是从状态 θ_i 到 θ_j 转移的期望值数目, 由此可以导出 Baum-Welch 算法中著名的重估公式表述如下:

$$\overline{\pi} \xi_1(i) \quad (16)$$

$$\overline{a}_{ij} = \sum_{t=1}^{T-1} \xi_t(i, j) / \sum_{t=1}^{T-1} \xi_t(i) \quad (17)$$

$$\overline{b}_{jk} = \sum_{t=1}^{T-1} \xi_t(i, j) / \sum_{t=1}^{T-1} \xi_t(i) \quad (18)$$

$\text{且 } o_t = v_k$

因此训练过程为由重估公式求得一组新参数 $\overline{\pi}, \overline{a}_{ij}, \overline{b}_{jk}$, 获得新模型 $\overline{\lambda} = (\overline{\pi}, \overline{A}, \overline{B})$, 可以证明 $P(O|\overline{\lambda}) > P(O|\lambda)$ 即重估公式所得 $\overline{\lambda}$ 能更好表示观测值序列 O 。重复此过程逐步改进模型参数, 直到 $P(O|\overline{\lambda})$ 满足收敛条件, 此时 $\overline{\lambda}$ 即为训练好的模型。

2 信号的特征矢量构成

2.1 全特征矢量构成

在时域、频域提取信号的全部统计特征, 在时频域通

过小波包分解提取不同频段的能量和能量熵特征,构成信号的特征矢量,这个特征矢量称为全特征矢量。本文对齿轮箱振动信号提取的时域统计特征有最大值、最小值、方根幅值、歪度、峭度、峭度指标、均值、绝对值、峰值、峰峰值、波形指标、峰值指标、方差、有效值、脉冲指标和裕度指标,共 16 个特征;频域统计特征有频谱总值和幅值谱熵,共 2 个特征。计算公式如式(19)和(20)所示。

$$F_1 = \frac{1}{M} \sum_{m=0}^{M-1} p(m) \quad (19)$$

$$F_2 = \frac{p(m)}{F_1} \cdot \lg M^{\frac{p(m)}{F_1}} \quad (20)$$

其中 $p(m)$ 为频谱幅值; $M(m=1, 2, 3, \dots, M)$ 为频谱线个数。

小波分解具有多分辨率分析能力,可将原始信号分解到不同频段,每个频段振动信号均含原始信号在该频段内的特征信息。对原始信号进行小波包分解,计算个频段信号的能量和能量熵,各频段信号能量计算如式(21)~(23)所示,能量熵计算如式(24)所示。

$$E_j = \sum_{k=1}^N x_k^2 \quad j=1, 2, 3, \dots, J \quad (21)$$

式中: x_k 为重构后信号; N 为重构信号长度; J 为原始信号分解频段个数。

总能量为:

$$E = \sum_{j=1}^J E_j \quad (22)$$

归一化后小波能量特征参数可表示:

$$\mathbf{W} = \left[\frac{E_1}{E}, \frac{E_2}{E}, \frac{E_3}{E}, \dots, \frac{E_J}{E} \right] \quad (23)$$

能量熵计算为:

$$ese_i = - \sum_{j=1}^N \left[\frac{|c_j|}{\sum_{j=1}^N |c_j|} \times \lg \left(\frac{|c_j|}{\sum_{j=1}^N |c_j|} \right) \right] / \lg N \quad (24)$$

式中: c_j 为小波包系数; i 为第 i 个小波包; N 为第 i 个小波包系数的个数;

由时域特征、频域特征和小波包分解后各个频段特征构成信号的全特征矢量。

2.2 全特征矢量构成

PCA 技术可以将描述信息的高维全特征矢量降为几个综合的指标,并保证损失信息量很少。PCA 分解后的主分量均为原始变量经线性组合变换得到的,各个主分量之间互不相关,使得主分量在表达信息方面的性能优于原始变量。

1) PCA 基本理论

设有 N 个样本,每个样本提取 P 个特征参量,则构成一个 $N \times P$ 的特征参量矩阵,该特征参量矩阵简记为:

$$\mathbf{X} = (X_1, X_2, X_3, \dots, X_N)^T$$

对于第 K 个样本的 P 个特征参量表示为:

$$\mathbf{X}_K = (x_{K1}, x_{K2}, x_{K3}, \dots, x_{KP}) \quad K=1, 2, \dots, N \quad (25)$$

特征参量矩阵为:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots \\ x_{21} & x_{22} & \dots \\ \vdots & \vdots & \dots \\ x_{N1} & x_{N2} & \dots \end{bmatrix} \quad (26)$$

对 p 维空间的任一向量 $\mathbf{X}_K$, 进行线性变换为:

$$\mathbf{Z}_K = \mathbf{A} \times \mathbf{X}_K \quad (27)$$

式中: $\mathbf{A}$ 为称线性变换的系数矩阵。

$\mathbf{Z}_K$ 的分量形式为:

$$\begin{cases} Z_{K1} = a_{11}x_{K1} + a_{12}x_{K2} + \dots + a_{1P}x_{KP} \\ Z_{K2} = a_{21}x_{K1} + a_{22}x_{K2} + \dots + a_{2P}x_{KP} \\ \vdots \\ Z_{KP} = a_{P1}x_{K1} + a_{P2}x_{K2} + \dots + a_{PP}x_{KP} \end{cases} \quad (28)$$

式中: $x_{K1}, x_{K2}, x_{K3}, \dots, x_{KP}$ 为第 K 个样本的 P 个原始分量; $Z_{K1}, Z_{K2}, Z_{K3}, \dots, Z_{KP}$ 为经线变换处理后的各个主分量, Z_{K1}, Z_{K2}, Z_{K3} 分别第 K 个样本的为第 1、第 2、第 3 主分量,以此类推。

2) PCA 算法

利用 PCA 求解特征矢量主成分的过程是求解线性变换系数矩阵 $\mathbf{A}$ 的过程。计算步骤如下:

(1) 计算特征参量矩阵 $\mathbf{X}$ 的协方差矩阵 $\mathbf{C}$ 。

$$\mathbf{C} = \frac{1}{P} (\mathbf{X} - \bar{\mathbf{X}}) (\mathbf{X} - \bar{\mathbf{X}})^T \quad (29)$$

式中: $\bar{\mathbf{X}}$ 是由 X 行向量均值构成的一个 $N \times P$ 矩阵,第 K 行的 P 个元素的数值计算如下:

$$\bar{x}_m = \frac{1}{N} \sum_{k=1}^N x_{km}, \quad m=1, 2, 3, \dots, P \quad (30)$$

协方差矩阵 $\mathbf{C}$ 是一个 $P \times P$ 的对称矩阵,其第 i 行、第 j 列元素值为:

$$c_{ij} = \frac{1}{P} \sum_{k=1}^N (x_{ik} - \bar{x}_i)^T (x_{kj} - \bar{x}_j) \quad (31)$$

$i, j=1, 2, \dots, P$

(2) 计算协方差矩阵 $\mathbf{C}$ 的特征值 $\lambda_i (i=1, 2, \dots, P)$ 和特征向量 $\mathbf{e}_{ij} (i, j=1, 2, \dots, P)$ 。其特征向量 $\mathbf{e}_{ij}$ 构成了线性变换的系数矩阵 $\mathbf{A}$ 。

(3) 对特征值的进行排序。由于特征向量对应的特征值与其对重构时主分量的贡献率呈正比关系,因此,可以通过对特征值的大小进行排列来选择第 1、第 2、 $\dots$ 、第 m 主分量 ($m < P$), 去除贡献率较小的 ($P-m$) 个主分量, 达到特征降维的目的。

如 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, 计算主分量累计贡献率为:

$$\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^P \lambda_i} \quad i=1,2,3,\cdots,P \tag{32}$$

若取主分量累积贡献率达 98% 以上的特征值包括 $\lambda_1, \lambda_2, \cdots, \lambda_m (m < p)$, 则所对应线性变换的 $Z_1, Z_2, \cdots, Z_m$ 分别称为第 1, 第 2, $\cdots$, 第 m 个主成分。由 m 个主分量组成的特征矢量称为主分量特征矢量。

3 基于特征矢量和 HMM 的齿轮箱故障诊断方法

本文分别采用全特征矢量—HMM 和主分量特征矢量—HMM 方法对齿轮箱的故障进行诊断。故障诊断流程如图 3 所示, 步骤如下。

- 1) 提取振动信号混合域 34 个特征参量, 构成全特征参量集;
- 2) 对混合域 34 维特征参量进行 PCA 降维处理, 选取吸收信息总量 98% 的主分量组成新的特征集;
- 3) 使用 Lloyds 算法分别对全特征参量和 PCA 降维后的主分量进行标量量化处理;
- 4) 使用 Baum-Welch 算法训练 HMM, 建立 HMM 模型库;
- 5) 将待诊断信号按照步骤 1) ~ 3) 提取全特征矢量和主分量特征矢量, 输入到训练好的 HMM 库中, 使用前-后向算法计算模型输出的对数似然概率, 似然概率最大的模型对应待诊断齿轮箱的故障模式。

4 齿轮箱故障诊断实验

4.1 实验平台介绍

为了验证本文采用方法的有效性, 在齿轮箱故障实验台上分别采集齿轮正常、齿轮点蚀和齿轮断齿情况下的振动信号, 该齿轮箱是一个二级减速箱, 齿轮箱的结构如图 4 所示。齿轮箱内齿轮均为圆柱齿轮, 其中输入轴上的齿轮齿数为 $Z_1 = 26$, 中间轴上双联大齿轮齿数为 $Z_2 = 64$, 中间轴上三联齿轮齿数为 $Z_3 = 40$, 输出轴齿数为 $Z_4 = 85$ 。三联齿轮从左到右分别是正常齿轮, 点蚀齿轮和断齿齿轮, 实验过程中, 通过换挡, 可实现齿轮箱在齿轮正常、齿轮点蚀故障和断齿故障 3 种状态下工作。

实验中采用的 3 个传感器, 都是 PCB 公司的 M601A01 型加速度传感器, 两个传感器分别安装在中间轴左、右轴承盖上, 另一个安装在输出轴的轴承盖上。分别开展齿轮正常、齿轮点蚀和齿轮断齿 3 种状态下振动实验。实验中输入轴的转速为 900 r/min, 每个通道的采

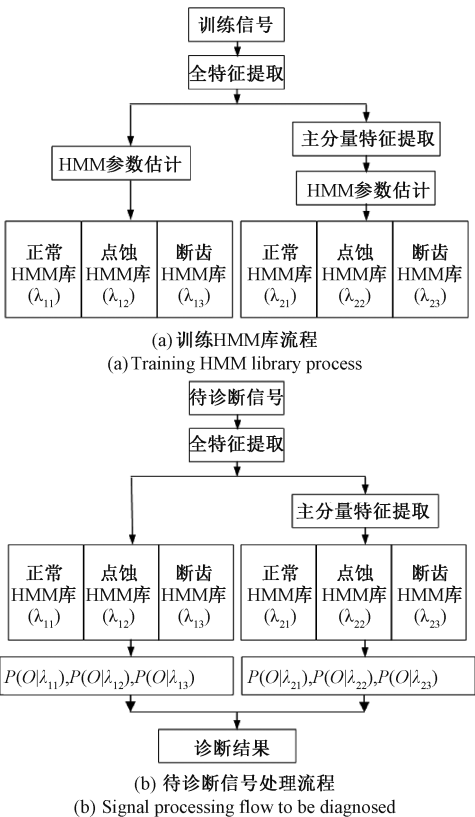


图 3 特征矢量-HMM 齿轮箱故障诊断流程
Fig. 3 Feature vector-HMM gearbox fault diagnosis process

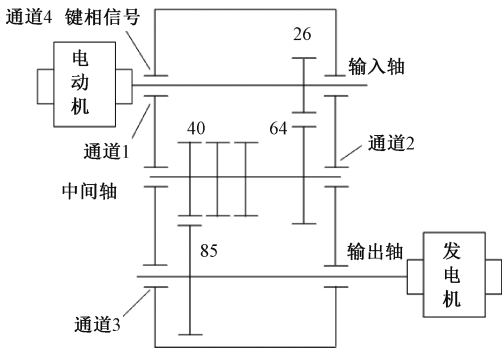


图 4 齿轮箱结构
Fig. 4 Gearbox structure

样频率为 16 kHz, 采样时间长度为 4 s。每种状态进行了 3 次, 图 5 所示为其中一次的测量结果。

4.2 齿轮箱振动信号特征提取和降维处理

在进行数据处理时, 先将每次测量数据分 16 段, 3 次测量数据共分为 48 段, 每段有 4 000 个采样数据, 分别对每段数据提取时域 16 个统计特征, 频域 2 个统计特征, 3 层小波包分解后 8 个频段的能量和能量熵 16 个特征, 混合域共计 34 个特征, 组成振动信号全特征矢量。3 类齿轮箱状态下一段测量数据的 34 个特征分布情况如

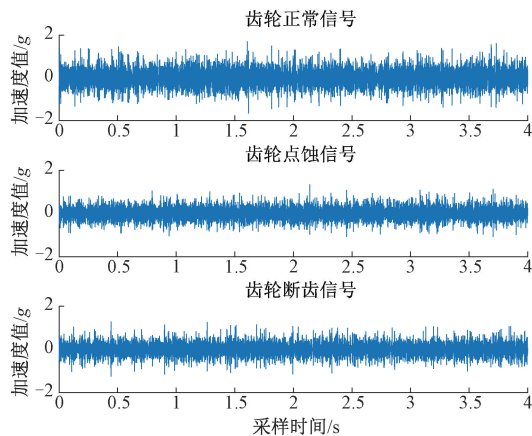


图 5 实验台采集的振动信号

Fig. 5 Vibration signal collected by the experimental platform

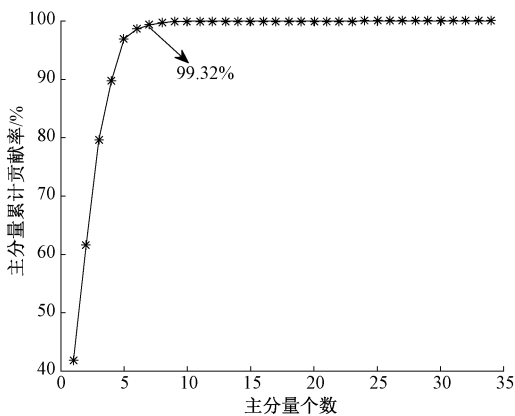


图 7 主分量成分累计

Fig. 7 Principal component accumulation

图 6 所示。

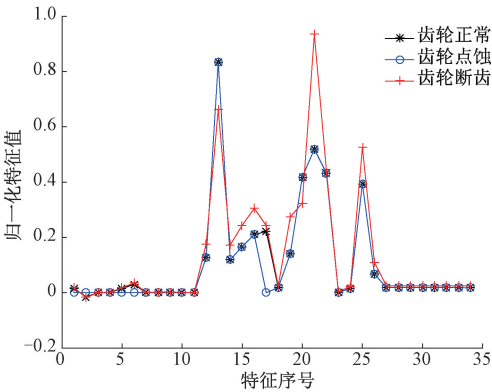


图 6 齿轮箱在 3 类状态下 34 个特征分布情况

Fig. 6 34 characteristics distribution of gearbox in three states

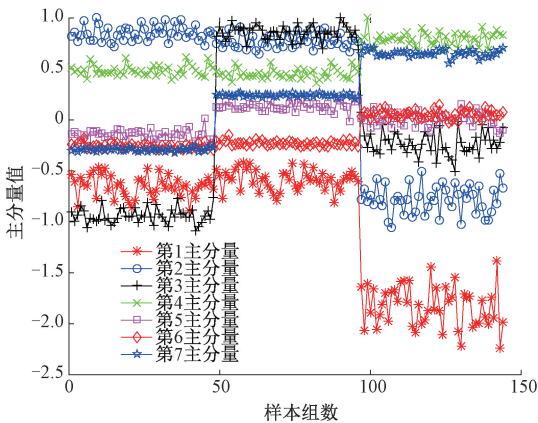


图 8 PCA 降维后 7 个特征分布

Fig. 8 Seven feature distributions after PCA dimensionality reduction

使用 PCA 对混合域的全特征参量集进行降维处理,图 7 所示为主分量成分累计分布,从图 7 可以看出前 7 个主分量累计信息量已超过 98%,所以取前 7 个主分量成分,将混合域的 34 维全特征矢量降到 7 维的主分量特征矢量。图 8 所示为 PCA 降维后的前 7 主分量分布情况,其中前 48 个样本为齿轮正常主分量分布,第 49~96 个样本是齿轮点蚀主分量分布,第 97~144 样本是齿轮断齿主分量分布。PCA 技术是正交变换,每两个主分量互不相关,之间的相关系数为 0,从而不存在每个主分量信息重叠现象。

4.3 训练 HMM 库

分别用混合域的全特征矢量和 PCA 降维后的 7 个 PCA 特征矢量训练 HMM,训练组用 25 个样本,测试组用全部的 48 个样本。

训练 HMM 库的迭代过程如图 9 所示,图中横坐标表示训练迭代次数,纵坐标表示在模型 $\lambda = (\pi, A, B)$ 下,

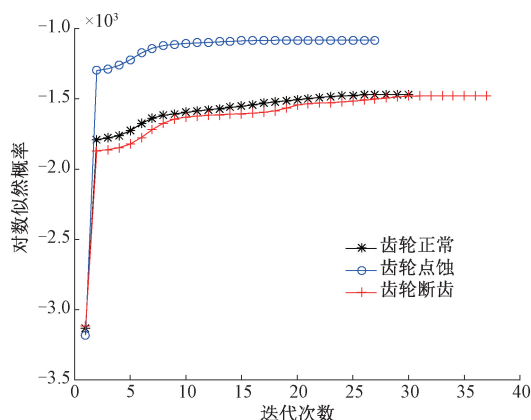
Baum-Welch 算法计算的训练样本值的输出概率,表示为 $P(O|\lambda)$ 。Baum-Welch 算法通过递归运算使 $P(O|\lambda)$ 达到局部最大。

由图 9 可以看出 HMM 经过几十步的迭代都达到了收敛误差(本文设置为 0.000 01),说明 HMM 具有很快的学习性能。比较两种方案训练 HMM 库的收敛速度,PCA 特征矢量训练正常 HMM 库、点蚀 HMM 库和断齿 HMM 库,训练速度均优于全特征训练速度。

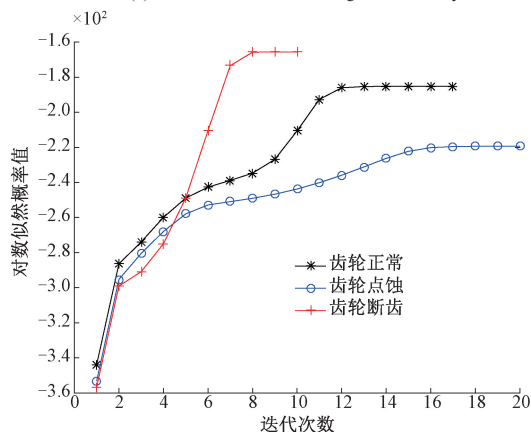
4.4 HMM 库的齿轮箱故障识别

为了判断两类 HMM 库对故障分类的准确率,将不同故障样本经特征提取、标量量化后输入给 HMM 库,用前向-后向算法计算 HMM 的输出概率,输出概率最大的 HMM,输入样本就为该模型的类别。

提取 48 组齿轮正常状态下振动信号时域、频域和小波包分解后每个包的特征,构成全特征矢量,经标量量化后输入给由全特征矢量训练的 3 类 HMM 库,诊断结果如



(a) 全特征矢量训练HMM库
(a) Full feature vector training HMM library



(b) PCA特征矢量训练HMM库
(b) PCA feature vector training HMM library

图 9 训练 HMM 迭代过程

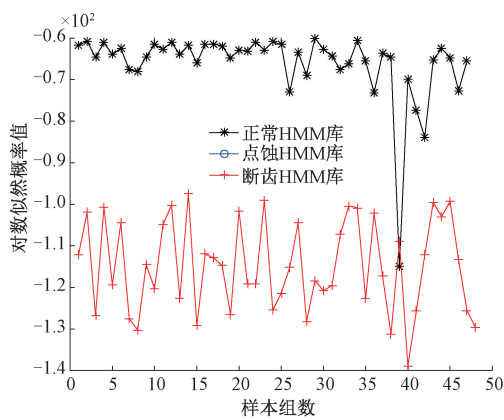
Fig. 9 Training HMM iterative process

图 10(a) 所示。同理, 分别对 48 组齿轮点蚀、齿轮断齿故障进行诊断, 诊断结果如图 10(b)、(c) 所示。

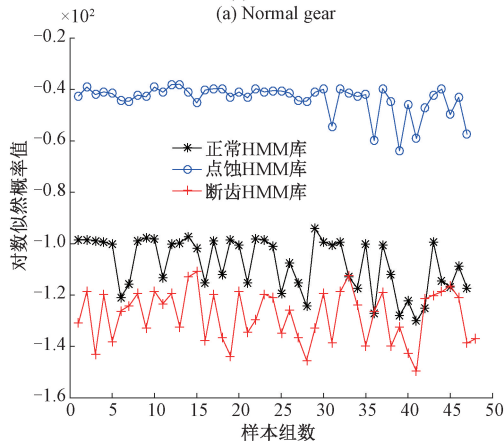
进一步对齿轮正常、齿轮点蚀和齿轮断齿的 48 组样本的全特征矢量进行 PCA 降维处理, 构成主分量特征矢量, 然后标量化后分别输入给由主分量特征矢量训练的 3 类 HMM 库, 诊断结果如图 11(a)~(c) 所示。

由图 10 分析全特征-HMM 库分类情况, 对齿轮正常的 48 个样本进行分类, 第 39 个样本分类错误, 其准确率为 97.9%, 对齿轮点蚀的 48 个样本进行分类, 分类全部正确, 其准确率为 100%, 对齿轮断齿的 48 个样本进行分类, 第 46 个样本分类错误, 其准确率为 97.9%。由图 11 分析 PCA 特征-HMM 库分类情况, 对齿轮正常的 48 个样本进行分类, 第 45 个样本分类错误, 其准确率为 97.9%, 对齿轮点蚀的 48 个样本进行分类, 全部正确, 其准确率为 100%, 对齿轮断齿的 48 个样本进行分类, 全部正确, 其准确率为 100%。

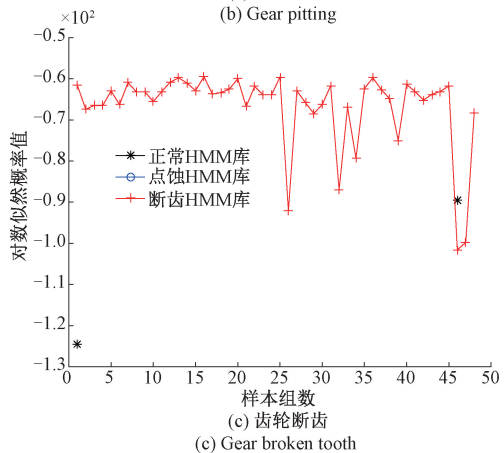
测试用全特征矢量和 PCA 特征矢量训练 HMM 和对样本进行分类程序运行时间, 全特征矢量完成 HMM 训



(a) 齿轮正常
(a) Normal gear



(b) 齿轮点蚀
(b) Gear pitting



(c) Gear broken tooth

图 10 全特征-HMM 诊断结果

Fig. 10 Full characteristics-HMM diagnosis results

练和 HMM 分类用时 22.238 s, PCA 特征矢量完成 HMM 训练和 HMM 分类用时 4.879 s。

比对两类 HMM 库, 在齿轮箱故障分类中的准确率, PCA 特征矢量训练的 HMM 在故障分类中准确率高于全特征矢量训练的 HMM; 从分类的速度上进行比较, 主分量特征-HMM 分类速度高于全特征-HMM 很多, 这在复杂的机械系统故障诊断和分类中是极其重要的。

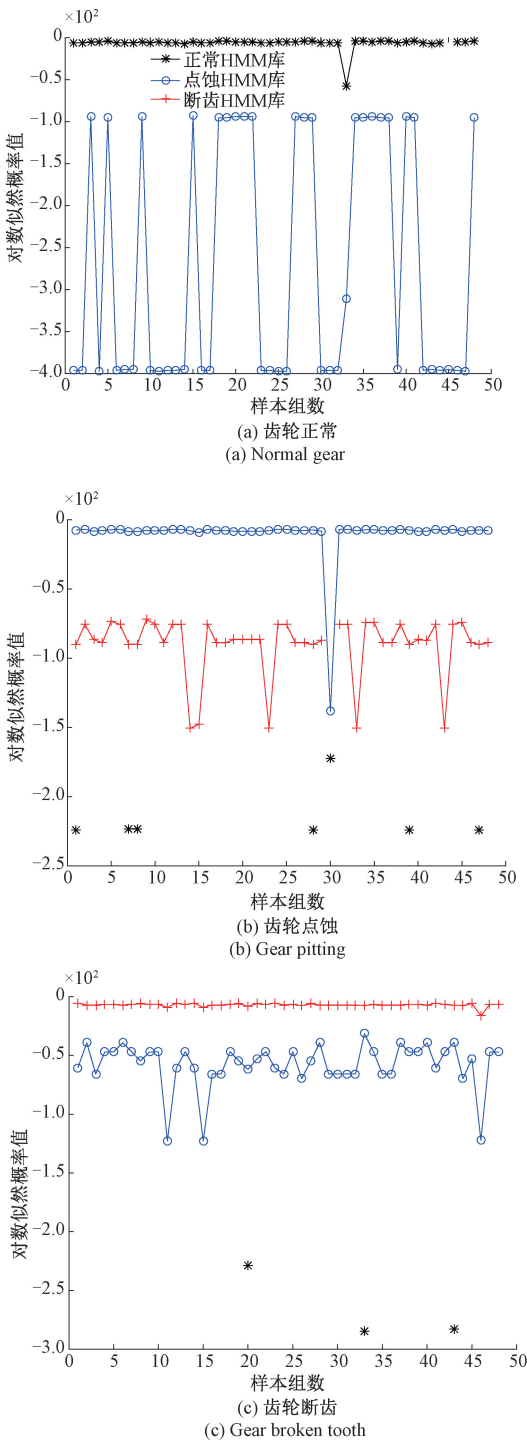


图 11 PCA 特征-HMM 诊断结果

Fig. 11 PCA characteristics-HMM diagnosis results

5 结 论

本文分析了 HMM 理论和 PCA 技术,将该理论和技术应用于对齿轮箱振动信号的处理中,建立了全特征-

HMM 库和 PCA 特征-HMM 库,开展了齿轮箱振动测试实验,通过对实验数据处理,验证了 HMM 理论和 PCA 技术在齿轮箱故障诊断和分类中的有效性,比较了两种 HMM 库在故障诊断和分类中的准确率以及数据处理速度,结果表明,PCA 特征-HMM 故障分类的准确率和诊断速度均优于全特征-HMM。本文的工作对机械故障诊断具有重要的实用价值。

参考文献

[1] 姜干新. 基于 HMM 的分布式语音识别系统的研究与应用[D]. 杭州:浙江大学,2010.
JIANG G X. Research on and application of HMM-based distributed speech recognition system [D]. Hangzhou: Zhejiang University,2010.
[2] 侯一民,周慧琼,王政一. 深度学习在语音识别中的研究进展综述[J]. 计算机应用研究,2017,34(8):2241-2246.
HOUY M,ZHOU H Q,WANG ZH Y. Overview of speech recognition based on deep learning [J]. Application Research of Computers,2017,34(8):2241-2246.
[3] MOHAMED A, DAHL G, HINTON G. Acoustic modeling using deep belief networks [J]. IEEE Transaction on Audio, Speech, and Language Processing,2012,20(1):14-22.
[4] BAUM L E, PETRIE T. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov processes [J]. Annals of Mathematical Statistics, 1970, 41:164-171.
[5] BAUM L E. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov process[J]. Inequalities, 1972(3):1-8.
[6] RABINER L R. Tutorial on hidden Markov models and selected applications in speech recognition [J]. Proceedings of the IEEE, 1989, 77(2):257-286.
[7] 盛敏,刘双庆,王婕,等. 基于 GMM-HMM 模型的智能下肢假肢运动意图识别[J]. 仪器仪表学报,2019,40(5):169-178.
SHENG M, LIU SH Q, WANG J, et al. Motion intent recognition of intelligent lower limb prosthesis based on GMM-HMM [J]. Chinese Journal of Scientific Instrument, 2019,40(5):169-178.
[8] 胡一帆,胡友彬,李骞,等. 基于视频监控的人脸检测跟踪识别系统研究[J]. 计算机工程与应用,2016,52(21):1-7,35.
HU Y F, HU Y B, LI Q, et al. Research on face detection, tracking and recognition system based on video surveillance[J]. Computer Engineering and Applications,

- 2019,40(5):169-178.
- [9] GUO X L, YANG T T. Gesture recognition based on HMM-FNN model using a Kinect [J]. Journal on Multimodal User Interfaces, 2016,11(1):1-7.
- [10] 耿庆田,赵宏伟. 基于分形维数和隐马尔科夫特征的车牌识别[J]. 光学精密工程, 2013, 21(12): 3219-3204.
- GENG Q T, ZHAO H W. License plate recognition based on fractal and hidden Markov feature [J]. Optics and Precision Engineering, 2013,21(12):3219-3204.
- [11] 潘月浩,宋执环,杜往泽,等. 适用于老年公寓的动作识别设计方法[J]. 电子测量与仪器学报, 2017, 31(1):29-35.
- PAN Y H, SONG ZH H, DU W Z, et al. Design method for action recognition applied to senile apartment [J]. Journal of Electronic Measurement and Instrumentation, 2017,31(1):29-35.
- [12] HECK L P, MCCLELLAN J H. Mechanical system monitoring using hidden Markov models[C] International Conference on Acoustics, Speech, and Signal Processing, IEEE, 2002:1697-1700.
- [13] BUNKS C, MCCARTHY D, AL-ANI T. Condition-based maintenance of machines using hidden Markov models [J]. Mechanical Systems and Signal Processing, 2000, 14(4): 597-612.
- [14] 童进,吴昭同,严拱标. 大型旋转机械升降速过程故障诊断 HMM-AR 方法研究[J]. 振动与冲击, 1999, 18(2):79-80,100.
- TONG J, WU ZH T, YAN G B. Study on HMM-AR method in fault diagnosing in speeding up or down process of large rotating machinery [J]. Journal of Vibration and Shock, 1999,18(2):79-80,100.
- [15] 冯长建,丁启全,吴昭同. 混合 SOM 和 HMM 方法在旋转机械升速全过程故障诊断中的应用[J]. 中国机械工程, 2002,13(2):1711-1714.
- FENG CH J, DING Q Q, WU ZH T. Hybrid SOM& HMM method in fault diagnosis applications of rotating machinery overall speed ascend& descend process [J]. China Mechanical Engineering, 2002,13(2):1711-1714.
- [16] 张西宁,雷威,杨雨薇,等. 采用自适应基因粒子群算法优化隐马尔科夫模型的方法及应用[J]. 西安交通大学学报, 2018,52(8):1-8.
- ZHANG X N, LEI W, YANG Y W, et al. Adaptive genetic particle swarm algorithm for optimization hidden Markov models with application [J]. Journal of Xi'an Jiaotong University, 2018,52(8):1-8.
- [17] DU J, HU Y, JIANG H. Boosted mixture learning of gaussian mixture hidden Markov models based on maximum likelihood for speech recognition [J]. IEEE Transactions on Audio, Speech, and Language, 2011, 19(7):2091-2100.

作者简介



杨秀芳(通信作者), 1989 年于上海理工大学获得学士学位, 1996 年于西安理工大学获得硕士学位, 现为西安理工大学副教授, 主要研究方向为光电检测与机械设备智能诊断。

E-mail: yxf5078@xaut.edu.cn

Yang Xiufang (Corresponding author)

received her B. Sc. degree from Shanghai University of Technology in 1989, M. Sc. degree from Xi'an University of Technology in 1996. Now she is an associate professor at Xi'an University of Technology. Her main research interests include photoelectric detection and intelligent diagnosis of mechanical equipment.