

DOI:10.3969/j.issn.1001-4551.2021.01.008

基于 EMDD 信息量和 KNP-SVDD 的 滚动轴承故障诊断研究*

陈宇晨,何毅斌*,戴乔森,刘 湘,贺苏逊
(武汉工程大学 机电工程学院,湖北 武汉 430205)

摘要:针对故障诊断中各类数据样本分布不均衡时诊断效果较差的问题,在支持向量机方法的基础上引入了支持向量描述方法,并对支持向量描述方法向多元判决扩展的方法,以及各种扩展方法的局限性进行了研究。对现有支持向量描述方法的扩展方法进行改进,提出了一种利用 K 相邻概率加权的多元判决支持向量数据描述方法;通过使用集合经验模态分解原信号,并计算了各本征模函数的信息量作为特征,利用第三方实验数据,对 K 相邻概率支持向量数据描述方法在各种故障类别的识别准确率进行了测试。研究表明:该方法可以有效地识别滚动轴承故障发生时的位置和严重程度;同时,通过与其他分类方法对比的方式,证明了该方法的优越性。

关键词:故障诊断;支持向量机;集合经验模态分解;本征模函数;K 相邻概率支持向量数据描述

中图分类号:TH133.33

文献标识码:A

文章编号:1001-4551(2021)01-0055-07

Fault diagnosis of rolling bearing based on EMDD information and KNP-SVDD

CHEN Yu-chen, HE Yi-bin, DAI Qiao-sen, LIU Xiang, HE Su-xun
(Mechanical and Electrical Engineering, Wuhan Institute of Technology, Wuhan 430205, China)

Abstract: Aiming at the problem of poor diagnosis effect when all kinds of data samples were unevenly distributed in fault diagnosis, support vector data description (SVDD) was proposed based on support vector machine (SVM), the extension of SVDD to multiple decision and the limitations of various extension methods were also studied. A multi-decision SVDD weighted by K-neighbor probability was proposed. The ensemble empirical mode decomposition (EEMD) was used to decompose the original signal, and the information content of each intrinsic mode function (IMF) was calculated and taken as a characteristic. The third-party experimental data were used to test the identification accuracy of k-neighbor probability support vector data description (KNP-SVDD) method in various fault categories. The results indicate that the method can effectively identify the location and severity of the fault, and the superiority of the method is proved by comparing with other classification methods.

Key words: fault diagnosis; support vector machine (SVM); ensemble empirical mode decomposition (EEMD); intrinsic mode function (IMF); K-neighbor probability support vector data description (KNP-SVDD)

0 引 言

滚动轴承被广泛应用于各种机械中。能够及时地

对滚动轴承进行监控,检测其是否受损,对工业生产具有重要的意义^[1]。滚动轴承的故障检测主要包括信号采集、特征提取、故障分类 3 个步骤。由于信号采集

收稿日期:2019-12-30

基金项目:湖北省科技厅重大专项资助项目(2016AAA056);化工装备强化与本质安全湖北省重点实验室开放基金资助项目(2018KA01)

作者简介:陈宇晨(1995-),男,湖北荆门人,硕士研究生,主要从事可视化与仿真、图像识别方面的研究。E-mail:1121130406@qq.com

通信联系人:何毅斌,男,博士,教授。E-mail:775700062@qq.com

的过程中存在一些噪声,导致特征提取不准确,需要对采集的原始信号进行处理。确保得到有用的特征是对故障进行准确分类的重要保障^[2]。

在对滚动轴承的信号进行处理和特征提取中,常用的方法有傅立叶变换(FT)、小波变换(WT)以及经验模态分解(EMD)等。FT和WT需要事先设定基函数,对非平稳信号的处理效果不好;EMD则是根据自身的时间尺度进行分解,可以很好地用于非平稳的信号,但EMD具有端点效应和模态混叠等缺陷。为了解决EMD的不足,集成经验模态分解(EEMD)被提出,并广泛应用到故障诊断的信号处理和特征提取过程中^[3];故障分类常用的方法有人工神经网络(ANN)、支持向量机(SVM)等,ANN需要很大的数据量以及很长的训练时间,而SVM可用于小样本训练,训练速度快,对于线性和非线性数据都有很好的分类效果,因此SVM广泛被应用于故障诊断的故障分类过程中^[4-6]。

为提高SVM分类的准确率,得到最佳的分类效果,需要使用一些优化算法,为SVM计算更好的惩罚参数以及核参数。陈法法等^[7]提出了EEMD能量熵和优化LS-SVM结合的滚动轴承故障检测方法;何青等^[8]提出了EEMD能量熵和MFFOA-SVM结合的滚动轴承故障检测方法;姬盛飞等^[9]提出了AFSA-SVM的滚动轴承故障检测方法;梁治华等^[10]提出了EEMD能量熵和CS-SVM结合的滚动轴承故障检测方法。

上述方法均取得了不错的效果,但由于SVM的限制,这些方法并不适用于各类别的样本数量不均衡的情况^[11-13],这种不均衡导致分类结果偏向于样本数量较多的那一类。

为了解决以上问题,研究人员引入了K相邻概率支持向量数据描述(KNP-SVDD)方法。SVDD是TAXD等^[14,15]在结构风险最小化的SVM基础上提出的单类描述方法,但这种方法仅可以用来判断样本是否属于某类。为了将SVDD扩展到多分类,王涛等^[16]通过训练多个SVDD模型,并比较样本到每个SVDD模型的相对距离的大小来判断样本的种类,该方法仅考虑模型对样本的影响而未考虑样本附近点对其影响,对越靠近SVDD模型边缘的样本判决结果的可靠性越低。付文龙等^[17]利用K最近邻(KNN)和SVDD法组合为I-SVDD方法判断样本的类别,即对同时满足多个SVDD模型的样本使用KNN方法根据其附近样本的信息判断类别,该方法仅考虑样本附近点对其影响而未考虑模型的影响,在训练样本少或各类别不均衡的情况下分类效果不好。

在此基础上,笔者引入KNP-SVDD方法,这种方法综合考虑模型以及样本附近点对其的影响,既包含整体的信息又包含局部信息,最后通过第三方实验数据,对该方法的诊断效果进行验证。

1 集成经验模态分解与信息量特征提取

1.1 集成经验模式分解

集成经验模态分解(EEMD)建立在经验模态分解(EMD)的基础上,通过往原信号中加入满足分布的白噪声信号,利用这些噪声均值为0的特性,均衡原信号噪声的特性,有效抑制了EMD中的模态混叠等缺陷。

EEMD的主要过程如下:

(1) 在原始信号 x_{old} 中加入满足 $N(0, \sigma^2)$ 分布的白噪声信号 ε ,组成新的信号 x_{new} ,即:

$$x_{new}(t) = x_{old}(t) + \varepsilon \sim N(0, \sigma^2) \quad (1)$$

(2) 利用EMD方法,将信号 $x_{new}(t)$ 分解为组本征函数(IMF)与一组残余分量(RES)之和,即:

$$x_{new}(t) = \sum_{i=0}^n IMF_i + RES \quad (2)$$

上式中,随着的数值由1向 n 增大,对应的 IMF_i 的频率宽度越小;

(3) 重复(1 ~ 2) m 次,共得到 m 组不同的 $\{IMF_1 \cdots IMF_n, RES\}$,对 m 组求平均值:

$$\begin{aligned} IMF_1 &= \sum_{i=1}^m IMF_{1i}/m \\ IMF_2 &= \sum_{i=2}^m IMF_{2i}/m \\ &\cdots \\ RES &= \sum_{i=2}^m RES/m \end{aligned} \quad (3)$$

可见,经过平均得到的结果即为EEMD分解后的各频率段的信号。

1.2 信息量特征提取

对滚动轴承故障的检测主要是通过在不同位置设置传感器,测得其振动信号。笔者根据振动信号对轴承故障进行判断,通过EEMD将测得信号分解为不同频率段的IMF以及残余分量RES;特征提取就是从不同频率段的IMF提取出具有代表性的量^[18]。

本文选择不同频率段IMF的信息量作为特征,特征提取的主要步骤如下:

(1) 假设第 i 个IMF在 j 时刻的函数值为 $f_i(j)$,则第 i 个IMF的幅值能量 E_i 可以表示为:

$$E_i = \|f_i(1) \cdots f_i(n)\|_2^2 \quad (4)$$

(2) 该信号的总能量 E_{sum} 可以表示为:

$$E_{\text{sum}} = \sum_i E_i \quad (5)$$

(3) 对各频率段 IMF 的幅值能量归一化处理,即:

$$p_i = E_i / E_{\text{sum}} \quad (6)$$

(4) 计算各频率段 IMF 的幅值能量归一化处理后
的信息量:

$$H_i = -\log_2(p_i) \quad (7)$$

最后,将信息量 $\{H_1, H_2, \cdots, H_n\}$ 作为特征,输入到
分类算法中。

2 K 相邻概率支持向量数据描述

2.1 支持向量数据描述

支持向量数据描述(SVDD)是一种单类数据描述
算法^[19],它将目标类数据通过核方法映射到一个高维
空间,并在高维空间中,通过目标类训练样本描述出一
个最符合样本的超球体;由于是对某一种样本的描述,
其不会受到不平衡样本的影响。

设 $\mathbf{X}_1 = \{x_1, x_2, \cdots, x_n\}$ 为收集到的 n 组同类别的
样本,该类别外所有的样本集合可表示为 $\mathbf{X}_2 = \{x_{n+1},$
 $x_{n+2} \cdots x_n\}$ 。

此时,问题可表述为利用样本 $\mathbf{X}_1$ 求目标超球体的
原点 a 和半径,该目标被表示成一个二次约束二次规
划(QCQP)的问题,即:

$$\begin{aligned} \min r^2 + C \sum_{i=1}^n \xi_i \\ \text{st } \|\varphi(x_i) - a\|_2^2 \leq r^2 + \xi_i \\ \xi_i \geq 0 \end{aligned} \quad (8)$$

式中: C —误差惩罚项系数; ξ_i —松弛变量,允许部分
训练样本的不在超球体内的情况出现; $\varphi(\cdot)$ —一个将
向高维空间映射的函数,用于解决样本在当前空间超
球体描述会造成很大误差的情况。

为了求解 QCQP 问题,笔者构建拉格朗日函数如
下:

$$\begin{aligned} L(r, \xi, \lambda, v) = r^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n v_i \xi_i - \\ \sum_{i=1}^n \lambda_i (r^2 + \xi_i - \|\varphi(x_i) - a\|_2^2) = \\ r^2 (1 - \sum_{i=1}^n \lambda_i) + \sum_{i=1}^n \xi_i (C - \lambda_i - v_i) + \\ \sum_{i=1}^n \lambda_i (\|\varphi(x_i) - a\|_2^2) \end{aligned} \quad (9)$$

式中: λ, v —拉格朗日乘子,且 $\lambda, v \geq 0$ 。

对式(9)中半径、圆心 a 和松弛变量 ξ_i 求偏导,并

令结果为0,则有:

$$\frac{\partial L}{\partial r} = 0 \Rightarrow \sum_{i=1}^n \lambda_i = 1$$

$$\frac{\partial L}{\partial a} = 0 \Rightarrow a = \sum_{i=1}^n \lambda_i \varphi(x_i) \quad (10)$$

将式(10)代入式(9)中,则式(8)的对偶形式可
以表示为:

$$\begin{aligned} \max \sum_{i=1}^n \lambda_i (\varphi(x_i), \varphi(x_i)) - \\ \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (\varphi(x_i), \varphi(x_j)) \\ \text{st } \sum_{i=1}^n \lambda_i = 1 \\ 0 \leq \lambda_i \leq C \end{aligned} \quad (11)$$

上式中 $\varphi(\cdot)$, 的形式未知,但 $\varphi(x)$ 间的内积可以
通过核函数 $K(\cdot)$ 计算,核函数需满足的条件如下:

$$K(x_i, x_j) = (\varphi(x_i), \varphi(x_j)) \quad (12)$$

本文使用的核函数均为高斯核函数,即:

$$K(x_i, x_j) = \exp\left(-\frac{\|\varphi(x_i) - \varphi(x_j)\|_2^2}{2\sigma^2}\right) \quad (13)$$

式中: σ —核宽度参数。

通过求解上述约束优化问题,可计算出符合样本
的超球体中心坐标。

对于训练样本中的所有点,到圆心的距离需满足
以下3种条件:

$$\begin{aligned} \|\varphi(x_i) - a\|_2^2 < r^2 \Rightarrow \lambda_i = 0 \\ \|\varphi(x_i) - a\|_2^2 = r^2 \Rightarrow 0 < \lambda_i < C \\ \|\varphi(x_i) - a\|_2^2 > r^2 \Rightarrow \lambda_i = C \end{aligned} \quad (14)$$

式(14)中,满足 $\|\varphi(x_i) - a\|_2^2 = r^2$ 的样本被称为支
持向量。为求超球体半径 r ,仅需要找到一个满足 $0 <$
 $\lambda_m < C$ 的样本 x_m 。超球体半径平方 r^2 可以表示为:

$$\begin{aligned} r^2 = \|\varphi(x_m) - a\|_2^2 = K(x_m, x_m) - \\ 2 \sum_{i=1}^n \lambda_i K(x_m, x_i) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j K(x_i, x_j) \end{aligned} \quad (15)$$

对于新的样本 z 是否与训练样本属于同一类,可
以利用其到球心的距离与半径的大小进行判断,即:

$$\begin{aligned} \|\varphi(z) - a\|_2^2 - r^2 = K(z, z) - 2 \sum_{i=1}^n \lambda_i K(z, x_i) + \\ \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j K(x_i, x_j) - r^2 \end{aligned} \quad (16)$$

若上式的结果小于等于0,表示样本 z 与训练样本
为同类。

2.2 K 相邻概率支持向量数据描述

为了将 SVDD 扩展到多元分类,通常采用组合

SVDD 的方法。组合 SVDD 圆形描述图如图 1 所示。

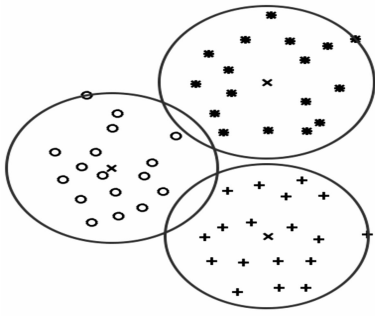


图 1 组合 SVDD 圆形描述图

图 1 中,使用 SVDD 方法对每一类别的数据做 1 次训练,组合多个圆形描述使之成为 1 个多类别分类器,再通过测试样本到球心的距离与半径的大小判断其类别^[20]。但 SVDD 只是针对未知结果的二值输出,若某一个测试样本落在多个圆形的相交区域,测试样本将同时被判定为多个类别。

为了处理以上这种情况,笔者利用 1 种基于概率的支持向量数据描述 (P-SVDD) 方法,将测试样本的球心的距离转换为属于该类别的概率^[21],即:

$$p_i = \frac{1}{1 + \exp(d_i - r_i)(1 + 1/d_i)} \quad (17)$$

式中: p_i —测试样本属于第 i 个超球体概率; d_i —测试样本到第 i 个超球体中心距离; r_i —第 i 个超球体半径。

当 d_i 逐渐增大时, p_i 逐渐减小;当 $d_i = 0$ 时, p_i 趋近 1;当 $d_i = r_i$ 时, $p_i = 0.5$;当 d_i 趋近于无穷时, p_i 趋近 0。通过计算测试样本属于每个超球体的概率,即可判定测试样本为概率最大的一类。

对于第 i 类 SVDD 中的惩罚项系数 C 和核宽度参数 δ ,可利用 1 组标签为 $Y = \{y_1, y_2 \cdots y_n\}$ 的样本,并通过公式进行优化,即:

$$\min_{C, \delta} - \left[\sum_{j=1}^n f(y_j) \log_2(p_j) + (1 - f(y_j)) \log_2(1 - p_j) \right]$$

$$f(y_j) = \begin{cases} 1 & y_j \text{ 属于第 } i \text{ 类} \\ 0 & y_j \text{ 不属于第 } i \text{ 类} \end{cases} \quad (18)$$

式(18)中,当计算的概率与真实的情况相差越远时,会赋予越大的惩罚;通过最小化惩罚,可以得到最佳的参数 C, δ ,从而得到最佳的 SVDD 模型。

将概率作为分类的判据,仅考虑了测试样本与超球体的半径和中心的信息,而没有考虑到测试样本附近点的信息,即只考虑到了总体的信息,未考虑到局部的信息。

信息的不足会导致对超球体边缘附近部分点的判决结果可信度不高,所以需要综合考虑这些因素的影响,更新测试样本的概率,具体的步骤如下:

(1) 设有训练样本 $X = \{x_1, x_2 \cdots x_n\}$ 及标签 $Y = \{y_1, y_2, \cdots y_n\}$ 和测试样本 z ,将 z 代到所有的训练并优化好的 SVDD 模型,得到的概率为 $p_1 \cdots p_n$;

(2) 若 $p_1 \cdots p_n$ 中没有值大于 0.5,则判决测试样本不属于任何一类;若 $p_1 \cdots p_n$ 中仅有一个值大于 0.5,则判决测试样本 z 属于 p_i 最大的那一类;若 $p_1 \cdots p_n$ 中有两个以上的值大于 0.5,利用下面的步骤更新大于 0.5 的概率 p_i ;

(3) 计算样本到训练样本 X 中所有样本的平方距离,即:

$$d_{zi}^2 = \|\varphi(z) - \varphi(x_i)\|_2^2 = K(z, z) - 2K(z, x_i) + K(x_i, x_i) \quad (19)$$

(4) 将所有的 d_{zi}^2 按照从大到小的顺序进行排列,排列后得到的集合为: $D_z^2 = \{d_{z1}^2, d_{z2}^2 \cdots d_{zn}^2\}$;

(5) 选择样本 z 的 k 个附近的样本,将 p 更新为:

$$p_{\text{new } i} = (1 - \omega) \times \frac{\sum_{i=1}^k f(y_i/d_{zi})}{\sum_{i=1}^k 1/d_{zi}} + \omega \times p_i \quad (20)$$

式中: $f(\cdot)$ —同式(18); ω —权重参数。

更新后的概率可以看作是原概率与其附近样本分布的加权。这种方法的可取之处是,既考虑了总体的信息,又考虑了局部的信息。

3 实验及结果分析

笔者采集了正常、内圈故障、外圈故障、滚动体故障 4 种状态的轴承振动信号;轴承的转速为 1 000 r/min,采样频率为 50 kHz,采集的时间为 5 s,根据条件可计算出当轴承每转动一圈时,采样的信号数量为 $60 \times 50\,000/1\,000 = 3\,000$ 个,采集 5 s 时共采样 250 000 个信号;以 3 000 个信号为一组,每种类别的故障信号共采集 80 组。

为了构造不平衡数据,笔者对滚动故障类仅选择 40 组,从每种类别中随机选择 60% 作为训练样本,40% 作为测试样本。

故障类型代号及抽样数量如表 1 所示。

表 1 故障类型代号及抽样数量

代号	故障类型	训练样本	测试样本
0	正常	48	32
1	内圈故障	48	32
2	外圈故障	48	32
3	滚动体故障	24	16

表 1 中,对 3 类故障样本数量选取较少,构造了一组不平衡的数据。笔者从每种类别中随机选择 60% 作为训练样本,40% 作为测试样本,对所有数据利用

EEMD 分解分别提取其特征。

笔者选择 EEMD 分解得到的前 7 阶 IMF 的信息量作为样本的特征,粒子群优化(PSO)计算惩罚项系数以及核宽度参数,取附近样本数为 10,权重参数为 0.5。

KNP-SVDD 单次诊断分类图如图 2 所示。

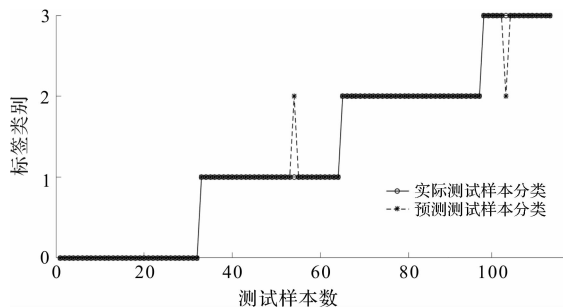


图 2 KNP-SVDD 单次诊断分类

图 2 中,利用 KNP-SVDD 方法,将表 1 测试样本中,16 个第 3 类故障中的 15 个进行了正确分类。

该方法对 112 个测试样本进行了测试,其中 107 个取得了正确的分类,准确率达到 98.214 3%。

为了比较不同方法之间的差异,笔者分别利用 I-SVDD 和相对距离 SVDD(RD-SVDD)、PSO-SVM 方法训练分类器,SVM 方法,用一对一方法扩展到多分类器。

I-SVDD 单次诊断分类图如图 3 所示。

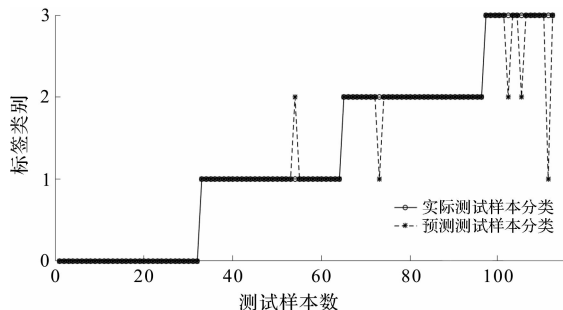


图 3 I-SVDD 单次诊断分类

图 3 中,利用 I-SVDD 方法,将表 1 测试样本中,16 个第 3 类故障中的 15 个进行了正确分类。

该方法对 112 个测试样本进行了测试,其中 107 个取得了正确的分类,准确率达到 95.535 7%。

RD-SVDD 单次诊断分类图如图 4 所示。

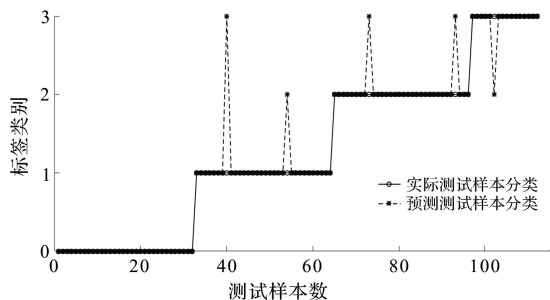


图 4 RD-SVDD 单次诊断分类图

图 4 中,利用 RD-SVDD 方法,将表 1 测试样本中,16 个第 3 类故障中的 15 个进行了正确分类。

该方法对 112 个测试样本进行了测试,其中 107 个取得了正确的分类,准确率达到 95.535 7%。

PSO-SVM 单次诊断分类图如图 5 所示。

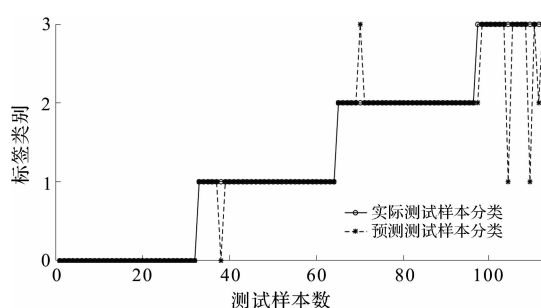


图 5 PSO-SVM 单次诊断分类图

图 5 中,利用 PSO-SVM 方法,将表 1 测试样本中,16 个第 3 类故障中的 11 个进行了正确分类。

该方法对 112 个测试样本进行了测试,其中 106 个取得了正确的分类,准确率达到 94.642 9%。

将 3 类的 16 个测试样本分对 11 个,并在所有测试样本上达到了 94.642 9% 的准确率。

图(3~5)中,由于 3 类故障类型的训练样本数据量较小,导致 PSO-SVM 方法对这类样本的误分率较高,而 I-SVDD 方法也会因这种不平衡,导致在该样本上的分类效果不好。利用 RD-SVDD 方法可以有效对这种不平衡类别的样本分类,上面 3 种方法的总体准确率大致相同。而无论是在不均衡数据的分类效果,还是总体分类准确率方面,KNP-SVDD 方法在这几种方法中均表现最好。

上面的结果仅代表 1 次诊断分类的结果,为了计算更准确的分类准确率以及对少数样本的误判个数,笔者重复上文的训练测试过程 20 次。

平均测试结果如表 2 所示。

表 2 平均测试结果

分类方法	准确率/%	3 类平均分对个数
KNP-SVDD	98.214 3	14.909 1
I-SVDD	95.454 5	12.909 1
RD-SVDD	95.454 5	15.181 8
PSO-SVM	95.211 0	11.727 3

表 2 中,在多次重复实验下,ISVDD、RD-SVD 与 PSO-SVM 这 3 种方法总体分类准确率差别不大,但 RD-SVDD 对数据量较少类别分类效果较好,证明了 SVDD 方法处理各类样本不均衡数据的有效性;而 PSO-SVM、I-SVDD 方法对数据量较少的类别分类效果

较差,说明在样本量小且不均衡的条件下,SVM 方法并不合适;而 KNN 方法仅考虑局部信息会导致 SVDD 方法失去部分算法本身的优势,KNP-SVDD 方法对比其他方法得到了最高的总体分类准确率,对数据量较少类别分类效果也较好,证明了该方法的可行性和有效性。

笔者利用美国凯斯西储大学轴承实验室标准数据库的轴承数据,再次对本文的方法进行验证。其中,选择电机负载为 0 hp,轴承的转速为 1 797 r/min,采样频率 12 kHz 的驱动端数据测试,故障类型分为 10 类。

故障类型代号及抽样数量如表 3 所示。

表 3 故障类型代号及抽样数量

代号	故障类型	训练样本	测试样本
0	正常	60	40
1	内圈损伤 0.018 cm	60	40
2	内圈损伤 0.036 cm	60	40
3	内圈损伤 0.053 cm	60	40
4	滚珠损伤 0.018 cm	60	40
5	滚珠损伤 0.036 cm	30	20
6	滚珠损伤 0.053 cm	60	40
7	外圈损伤 0.018 cm	60	40
8	外圈损伤 0.036 cm	30	20
9	外圈损伤 0.053 cm	60	40

表 3 中,每种类型 60% 作为训练样本,40% 作为测试样本,选择 EEMD 分解得到的前 6 阶 IMF 的信息量作为样本的特征,取附近样本数为 15,权重参数为 0.5。

不同方法分类准确率如表 4 所示。

表 4 不同方法分类准确率

分类方法	KNP-SVDD	I-SVDD	RD-SVDD	PSO-SVM
准确率/%	98.982 0	96.675 9	97.229 9	96.592 9

不同方法对每种故障分类准确率如图 6 所示。

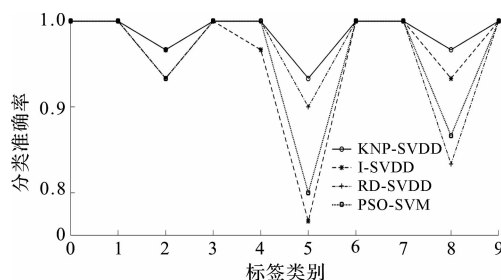


图 6 不同方法对每种故障分类准确率

表 4 和图 6 中,通过对比这 4 种方法可知:

(1) PSO-SVM 方法对第 5 类和第 8 类故障类型的分类效果较差,I-SVDD 方法对第 8 类故障类型的分类效果较好,但对第 5 类故障类型的分类效果较差;

(2) RD-SVDD 方法对第 5 类故障类型分类效果较好,对第 8 类故障类型的分类效果较差;

(3) KNP-SVDD 方法在全部测试样本上,以及每一种故障类别上的分类准确率均达到了最高,再一次验证了该方法的可行性和有效性。

4 结束语

本文基于 K 近邻(KNN)法和概率支持向量描述(P-SVDD)法,提出了 K 相邻概率支持向量数据描述(KNP-SVDD)法,用于提升各类数据分布不均匀的情况下的滚动轴承故障诊断的识别率;并通过第三方实验数据验证了在分布不均衡的数据方面,该方法可以取得较好的总体分类以及各类分类效果;

最后笔者将该方法与现有的一些支持向量机(SVM)方法、支持向量数据描述(SVDD)方法进行了比较,结果发现本文方法优于进入对比的各种方法,说明了该方法的可行性和有效性;同时,该方法对滚动轴承故障诊断的研究具有重要的参考价值。

参考文献(References):

- [1] 张超,陈建军,郭迅. 基于 EEMD 能量熵和支持向量机的齿轮故障诊断方法[J]. 中南大学学报,2012,43(3):932-939.
- [2] AI YT, GUAN J Y, FEI C W, et al. Fusion information entropy method of rlling bearing fault diagnosis based on n-dimensional characteristic parameter distance[J]. **Mechanical Systems&Signal Processing**,2017(88):123-136.
- [3] 黄艳庭,谷玉海,黄竞楠,等. 基于 EEMD 和自相关阈值降噪的滚动轴承故障诊断[J]. 组合机床与自动化加工技术,2019(10):78-82.
- [4] CHEN Ru-yi. Fault diagnosis of rolling bearing based on EEMD information entropy and improved SVM[C]. proceedings of the 38th China Control Conference, china:[s, n.],2019.
- [5] 何庆飞,陈小虎,姚春江,等. 基于最小二乘支持向量分类机的出轮泵故障诊断研究[J]. 流体机械,2019,47(8):32-36.
- [6] 张璐莹,李伟,姜智通,等. 基于主成分分析与 SVM 的碳纤维复合材料损伤声发射信号模式识别[J]. 压力容器

- 器,2020,37(7):59-68.
- [7] 陈法法,李 晁,陈保家,等. EEMD 能量熵与优化 LS-SVM 的滚动轴承故障诊断[J]. 组合机床与自动化加工技术,2016(12):71-75.
- [8] 何 青,褚东亮,毛新华. 基于 EEMD 和 MFFOA-SVM 滚动轴承故障诊断[J]. 中国机械工程,2016,27(9):1191-1197.
- [9] 姬盛飞,王丽君,吉南阳. 基于 AFSA-SVM 的滚动轴承故障诊断研究[J]. 组合机床与自动化加工技术,2019(1):115-117.
- [10] 梁治华,曹江涛,姬晓飞. 基于 EEMD 和 CS-SVM 的滚动轴承故障诊断研究[J]. 机电工程,2019,36(6):622-627.
- [11] 王 斐,房立清,赵玉龙,等. 基于 VMD 和 SVDD 的滚动轴承早期微弱故障检测和性能退化评估研究[J]. 振动与冲击,2019,38(22):224-230.
- [12] 张 媛,杜艳平,高振清,等. 基于 PSO-SVDD 正常域的无故障数据条件下轴承隐患监测[J]. 现代制造工程,2019(8):94-99.
- [13] TAO Xin-min, CHEN Wan-Hai, DU Bao-xiang, et al. A novel model of one-class bearing fault detection using SVDD and genetic algorithm [P]. Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on,2007.
- [14] TAX D, DUIN R. Support vector domain description[J]. **Pattern Recognition Letters**, 1999, 20 (11-13) : 1191-1199
- [15] TAX D M J, DUIN R P W. Support vector data description [J]. **Machine Learning**, 2004,54(1):45-66.
- [16] 王 涛,李艾华,王旭平,等. 基于 SVDD 与距离测度的齿轮泵故障诊断方法研究[J]. 振动与冲击,2013,32(11):62-65.
- [17] 付文龙,谭佳文,王 凯. 基于 VMD 散布熵与改进灰狼优化 SVDD 轴承半监督故障诊断研究[J]. 振动与冲击,2019,38(22):190-197.
- [18] WU Yu-Xi, WU Qing-Biao, ZHU Jia-Qi. Improved EEMD-based crude oil price forecasting using LSTM networks [J]. **Physica A: Statistical Mechanics and its Applications**,2019,516(27):114-124.
- [19] AZAMI M E, LARTIZIEN C, CANU S. Converting SVDD scores into probability estimates: Application to outlier detection[J]. **Neurocomputing**,2017,268(13),64-75.
- [20] 杨艳君,魏永合,王晶晶,等. 基于 LMD 和 SVDD 的滚动轴承健康状态评估[J]. 机械设计与制造,2019(5):163-166.
- [21] 杨 晨,王婕婷,李飞江,等. 基于概率的支持向量数据描述方法[J]. 计算机应用,2019,39(11):3134-3139.

[编辑:冯黎雅]

本文引用格式:

陈宇晨,何毅斌,戴乔森,等. 基于 EMDD 信息量和 KNP-SVDD 的滚动轴承故障诊断研究[J]. 机电工程,2021,38(1):55-61.

CHEN Yu-chen, HE Yi-bin, DAI Qiao-sen, et al. Fault diagnosis of rolling bearing based on EMDD information and KNP-SVDD[J]. *Journal of Mechanical & Electrical Engineering*, 2021,38(1):55-61.

《机电工程》杂志: <http://www.meem.com.cn>