

DOI: 10.19650/j.cnki.cjsi.J2006458

基于改进型 shapelets 算法的动车组轴箱 轴承故障诊断方法研究

宋志坤, 徐立成, 胡晓依, 任海星, 李 强

(北京交通大学机械与电子控制工程学院 北京 100044)

摘要: 现存的两种分别基于信号处理技术和大数据处理技术的滚动轴承故障诊断方法, 存在着过度依赖信号处理、模型复杂、可解释性弱等特点。针对传统故障诊断技术的不足, 本文将基于 shapelets 学习算法的时间序列分类方法引入故障诊断领域, 通过动车组轮对台架滚振实验建立了动车组轴箱轴承故障的非平衡数据集, 并基于 Dropout 思想对诊断模型进行了改进。实验结果表明, 该方法在保证故障诊断精确度的同时, 保留了 shapelets 作为“最具代表性的时间序列子序列”的强可解释性。同时, 基于 Dropout 的模型改进提升了模型的泛化性能, 在轴承故障数据的训练集和测试集上都取得了 100% 的诊断精度, 证明了基于 shapelets 的改进学习算法是一种可行的应用于动车组轴箱轴承故障诊断的方法。

关键词: 故障诊断; 滚动轴承; shapelets; 机器学习; 动车组

中图分类号: TH113.1 TH165.3 **文献标识码:** A **国家标准学科分类代码:** 460.4099

Research on fault diagnosis method of axle box bearing of EMU based on improved shapelets algorithm

Song Zhikun, Xu Licheng, Hu Xiaoyi, Ren Haixing, Li Qiang

(School of Mechanical, Electronic and Control Engineering, Beijing Jiatong University, Beijing 100044, China)

Abstract: The two currently existing fault diagnosis methods for rolling bearings based on signal processing technology and big data processing technology have the disadvantages of over-reliance on signal processing, complicated model and weak interpretability. Aiming at the shortcomings of traditional fault diagnosis technologies, this paper introduces the time series classification method based on shapelets learning algorithm into the field of fault diagnosis, and establishes the unbalanced data set of the faults of the EMU axle box bearing through the EMU wheelset bench rolling vibration experiment. The diagnosis model is improved based on the idea of Dropout. Experiment results show that the method guarantees the accuracy of fault diagnosis while retaining the strong interpretability of the shapelets as “the most representative time series subsequences”. At the same time, the improvement of the model based on Dropout improves the generalization performance of the model. The diagnostic accuracy of 100% on the training set and test set of bearing fault data was achieved, which proves that the improved learning algorithm based on shapelets is a feasible method applied to the fault diagnosis of axle box bearing of electric multiple unit.

Keywords: fault diagnosis; rolling bearing; shapelets; machine learning; electric multiple unit

0 引 言

走行部位于动车组每节车厢的两端,对动车组的安全平稳运行起到重要作用。随着动车组速度的提升,走行部部件在高速重载工况下的工作条件日益严峻,故障

时有发生。其中,轴箱滚动轴承作为动车组走行部最关键的旋转部位,对其早期故障的及时发现和识别具有重要的工程实际意义。传统的故障诊断方法主要是基于信号处理方法的诊断,随着计算机技术的发展,将信号处理方法和大数据处理方法相关联,已经成为了当前该方法的主要发展趋势,如将迭代小波变换和稀疏滤波相结合

的故障特征提取^[1],将经验模态分解后得到的本征模态函数与奇异值分解相结合等^[2]。

现存的滚动轴承的故障诊断主要可以分为 3 类,1) 基于数学模型的诊断,2) 基于经验知识的诊断,3) 基于数据驱动的诊断。基于数据驱动的智能诊断即利用大数据处理方法(如机器学习算法等)对故障的大数据实现智能分类。其诊断流程为先对不同工况下的轴承进行信号采集得到振动信号,再用信号处理方法对振动信号进行特征提取得到其时频特征,最后通过智能诊断方案对故障实现分类并输出评价和决策。

基于数据驱动的智能诊断方法通过对经典机器学习算法的改进,有针对性地实现机械故障的诊断。如:刘长良等^[3]将变模态分解和奇异值分解的特征提取方法与标准模糊 C 均值聚类(fuzzy C means clustering, FCM)相结合,采用海明贴近度对轴承故障进行分类。陈超等^[4]提出了一种应用于数据量不足情况的迁移学习策略,基于增强型最小二乘支持向量机(support vector machine, SVM),在目标函数和约束条件中分别增加辅助集的惩罚系数和约束条件,实验结果表明,该方法可在振动数据较少的条件下提升模型的故障诊断性能。曲建岭等^[5]提出了一种基于端到端的自适应一维卷积神经网络的故障诊断方法,利用深度网络结构实现对振动信号特征的自适应层级化提取,并利用 Softmax 分类器输出诊断结果。Feng 等^[6]提出的基于卷积神经网络诊断方法,将改进型正则化卷积核和权重 Softmax 损失函数引入诊断过程,并对每个隐藏层实现了可视化,提出了可将卷积核视为自适应滤波器的重要观点。

这类基于数据驱动的故障诊断方法计算速度快,诊断精度高,但将一个工程实际问题转化为一个数学模型的建立问题,使得故障诊断的可解释较差,所提取的特征不具备物理意义。基于 shapelets 的时间序列分类方法是近年来提出的一种基于时间序列相似性度量的分类方法^[7],将其引入故障诊断领域,可以在保证诊断精度的同时,提升诊断方法的可解释性和实际意义。近年来,基于 shapelets 相关概念的方法不断涌现,如王志海^[8]提出的基于 shapelets 的懒惰式时间序列分类模型,原继东^[9]采用的基于智能缓存计算重用技术的逻辑 shapelets 时间序列分类方法等,都是 shapelets“最具有代表性的时间序列子序列”的变种应用,并在某些特定的数据集上证明了该分类方法的可行性。

1 shapelets 发现算法

shapelets 是由文献[7]提出来的一种信号分类方法,该方法首先提取信号数据集中最具有代表性的子信号(即 shapelets),然后计算 shapelets 与各条信号之间的相

似度并进行比较,从而实现信号的分类。shapelets 的相关定义与符号如表 1 所示。

表 1 shapelets 相关概念符号表
Table 1 Symbol table related to shapelets concepts

符号	意义
D	信号数据集
T, R	单条信号
S	子信号
$N, D $	信号的个数
subdist	长度不等的子信号距离
E	熵
I	信息增益
G	分裂间隔

对于一个时间序列数据集 $D = \{T_1, T_2, \dots, T_n\}$, 每条时间序列 T_i 具有相同的维度 m , 即 $T_i = \{t_{i1}, t_{i2}, \dots, t_{im}\}$ 。时间序列子序列 $s_{i,1}$ 是一条时间序列 T_i 中从位置 i 开始, 长度为 l 的连续子序列, 即 $s_{i,1} = \{t_i, t_{i+1}, \dots, t_{i+l-1}\}$ 。

对周期性时间序列进行相似度量时, 往往是利用计算两者之间的欧式距离来进行比较, 距离越小则相似程度越大; 距离越大则相似程度越小, 即对于两条长度同为 m 的时间序列 q 和 c , 其相似度可以表示为:

$$\text{dist}(q, c) = \sqrt{\sum_{i=1}^m (q_i - c_i)^2} \quad (1)$$

但对于滚动轴承的实际工况而言, 传感器对其加速度数据的采样可能会导致在分割时间序列时, 时间序列的起始位置无法正确对应轴承运转的起始位置。因此, 在计算时间序列相似度时, 往往是通过计算时间序列子序列之间的相似度。对于长度不一样的两条时间序列 q 和 c , 其长度分别为 $|q|$ 和 $|c|$, 且 $|q| < |c|$ 。本文的做法是将较短的时间序列在较长的时间序列上滑动, 计算时间序列 q 和所有长度为 $|q|$ 的时间序列 c 的子序列(表示为 $c_{|q|}$)之间的欧式距离(其他可用的时间序列相似性度量包括: 动态时间规整 DTW^[10]、模式距离^[11]、最长公共子序列 LCSS^[12]等), 并比较大小, 将其中的最小值定义为时间序列的子序列距离, 将其定义为 subdist, 即:

$$\text{subdist}(q, c) = \min(\text{dist}(q, c_{|q|})) \quad (2)$$

shapelets 就是一个由 T_i 的子序列 s 和一个距离阈值 t 所组成的元组 (s, t) 。对于一个分类问题, 寻找最优的 shapelets 的目的是使得候选 shapelets 与所有同类时间序列的子序列距离最小, 而与不同类的时间序列的子序列距离最大, 从而达到分类的目的。文献[7]中提出 shapelets 暴力发现算法, 通过将所有可能的 shapelets 作为候选 shapelets, 通过计算信息增益和分类间隔等指标来对 shapelets 的分类效果进行评价和排序, 最终挑选出

分类结果最好的 K 个 shapelets。

在实际计算过程中, shapelets 暴力发现算法是一种耗时极长的循环算法, 借助计算机的强大计算能力, 尽管可以保证提取出来的子信号一定来源于原信号, 但循环算法的耗时较长, 且从分类的结果来说, 选择来源于原信号子信号作为 shapelets 并不一定能够获得最好的分类效果。Grabocka 等^[13]提出的一种基于机器学习的 shapelets 算法, 有效地解决了上述问题。

以一个二分类问题为例, 若模型需要将 I 个时间序列进行分类, 则分类的实际结果 $Y \in \{0, 1\}^I$, 预测结果 \hat{Y}_i 可以表示为 I 个时间序列与该 shapelet 的相似度的线性叠加, 再经过 sigmoid 函数归一化表示为 $[0, 1]$ 区间内的概率 $\hat{P}_i = \sigma(\hat{Y}_i)$, 数值上越接近 1 则为类 1 的概率越大; 数值上越接近 0 则为类 1 的概率越小, 为类 0 的概率越大:

$$\hat{Y}_i = W_0 + \sum_{k=1}^K M_{i,k} W_k \quad (3)$$

定义损失函数来计算预测结果与实际结果之间的误差, 并将其作为目标函数 F :

$$L(Y, \hat{Y}) = -Y \ln \hat{Y} - (1 - Y) \ln(1 - \hat{Y}) \quad (4)$$

机器学习算法的核心是梯度下降算法。它将目标函数和参数优化两者联系起来, 由于机器学习模型的最终目的是最小化目标函数, 而梯度下降算法就是实现最小化目标函数的一种常见方式。以一个二次目标函数为例, 为找到目标函数的最小值, 即目标函数图像的最低点, 则需要沿着函数曲线下降的方向进行寻找, 直到找到曲线的最低点; 换言之, 寻找目标函数最小值的跟踪曲线上每一点的斜率一定是负数, 因此, 自变量的取值是朝着负梯度(负导数)的方向进行更新, 直到找到最小值点。通过人为地设置一个学习率来控制自变量每次更新的幅度。当目标函数对自变量的梯度无法小于 0, 即陷入最小值, 梯度为 0, 左右两侧的梯度都大于 0 时, 则目标函数的最小值点已被找到。如图 1 所示, 该图像演示了一个目标函数的梯度下降过程。

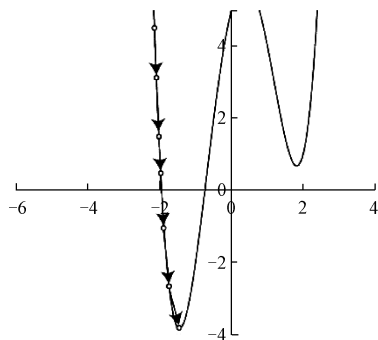


图 1 目标函数的梯度下降过程

Fig. 1 The gradient descent process of the objective function

由于离散信号为不连续的物理量, 因此很难直接计算损失函数对 shapelets 和超平面参数的梯度。文献[13]提出了一种 soft-min 函数来实现目标函数梯度的计算, 该方法原理较为复杂, 但效果十分显著, 在此不再赘述其原理。计算结果表明, 目标函数对第 k 个长度为 l 的 shapelets 的微分可以表示为:

$$\frac{\partial F_i}{\partial S_{k,l}} = \frac{\partial L(Y_i, \hat{Y}_i)}{\partial \hat{Y}_i} \frac{\partial \hat{Y}_i}{\partial \hat{M}_{i,k}} \sum_{j=1}^I \frac{\partial \hat{M}_{i,k}}{\partial D_{i,k,j}} \frac{\partial D_{i,k,j}}{\partial S_{k,l}} \quad (5)$$

对超平面参数 W_k 和 W_0 的微分可以表示为:

$$\frac{\partial F_i}{\partial W_k} = - (Y_i - \sigma(\hat{Y}_i)) \hat{M}_{i,k} \quad (6)$$

$$\frac{\partial \hat{Y}_i}{\partial W_0} = - (Y_i - \sigma(\hat{Y}_i)) \quad (7)$$

因此, 对于长度确定为 L 的 K 个 shapelets 分类问题, 可通过表 2 来实现 shapelets 和 W 的更新, 通过给定一个包含 I 个时间序列, 维度都是 Q 的时间序列数据集 T , 给定 shapelets 的个数 K , 确定的 shapelets 长度 L , 正则化项系数 λ_w , 学习率 η 和迭代次数 N , 即可利用随机梯度下降的原理和反向传播算法, 得到最优的 K 个 shapelets、超平面参数 W 和 W_0 。

表 2 shapelets 学习算法

Table 2 Shapelets learning algorithm

Learning Time Series Shapelets(T, K, L, η, N)
输入: 时间序列 T , shapelets 个数 K , 长度 L , 学习率 η 和迭代次数 N
输出: shapelets, 超平面参数 W 和 W_0
1: Initialize $S_{k,l}, W, W_0$
2: for iteration = 1, ..., N do
3: for $i = 1, \dots, I$ do
4: for $k = 1, \dots, K$ do
5: $W_k = W_k - \eta(\partial F_i / \partial W_k)$
6: for $l = 1, \dots, L$ do
7: $S_{k,l} = S_{k,l} - \eta(\partial F_i / \partial S_{k,l})$
8: end for
9: end for
10: $W_0 = W_0 - \eta(\partial F_i / \partial W_0)$
11: end for
12: end for
13: return S, W, W_0

上述过程为一针对确定 shapelets 长度的时间序列分类算法, 为保证 shapelets 的计算精度, 同时减少 shapelets 的计算耗时, 可以通过设定一个最小候选 shapelets 长度 L_{\min} , 所有候选的 shapelets 的长度可表示为集合 $\{L_{\min}, 2L_{\min}, \dots, \Delta L_{\min}\}$, 第 δ 类 shapelets 的长度可以表示为 δL_{\min} 。显然, 对于长度为 Q 的时间序列, 长度为 δL_{\min} 的

候选子序列个数可以表示为 $J(\delta) = Q - \delta L_{\min} + 1$ 。通过对表 2 算法添加一个 for 循环,即可计算得到不同长度的 shapelets。

算法的关键步骤之一是 shapelets 和超平面参数的初始化。超平面参数采用在 0 附近取随机值的方法,而初始化 shapelets 采用对所有长度为 L 的候选 shapelets 的 k-medoids 质心。

给定一个包含 I 个时间序列,维度都是 Q 的时间序列数据集 T ,首先从所有长度为 L 的子序列中随机选取 K 个初始化子序列,则每一条时间序列中包含了 $J = Q - L + 1$ 条长度为 L 的子序列,对 T 中共 $I \cdot J$ 个长度为 L 的子序列分别计算其与 K 个初始化子序列之间的相似度,将 $I \cdot J$ 个时间序列划分到与其相似度最高的初始化子序列的类中,形成 K 个聚类,在每个聚类中重新计算新的质心,具体做法为:计算该聚类中每一个序列到某一个序列的相似度之和,该相似度之和即选择该子序列作为该聚类质心的绝对误差,选择使得聚类绝对误差作为该聚类的质心。重复上述步骤直满足迭代次数要求或聚类绝对误差小于某阈值,迭代所得到的 K 个子序列即为所有长度为 L 的候选 shapelets 的质心,用这 K 个子序列作为初始化 shapelets 可在一定程度上加快模型的收敛。

基于 shapelets 学习算法的诊断模型的结构和 shapelets 更新过程如图 2 所示。

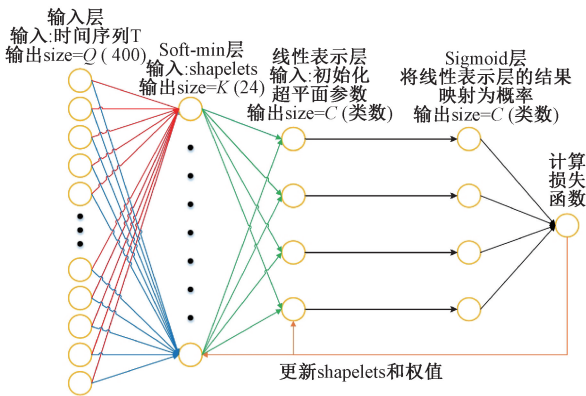


图 2 shapelets 学习算法结构

Fig. 2 Structure of the shapelets learning algorithm

2 动车组轴箱轴承故障数据采集实验

目前大量基于数据驱动的智能诊断方案面临的一个重要问题是:用于模型验证的数据集较少,美国凯斯西储大学数据集 (CWRU) 作为最常用的轴承故障数据集,其所包含的轴承工况十分有限。对于某一特定类型的轴承故障,无法对诊断模型进行有针对性的改进。为解决故障诊断方法对 CWRU 数据集的依赖,本文基于动车组轮

对台架滚振试验台进行故障数据采集试验,为故障诊断模型提供数据支撑。

该滚振试验台由中国铁道科学研究所和德国 RENK 试验系统公司联合设计,最高试验速度可达 500 km/h。试件的安装图如图 3 所示。

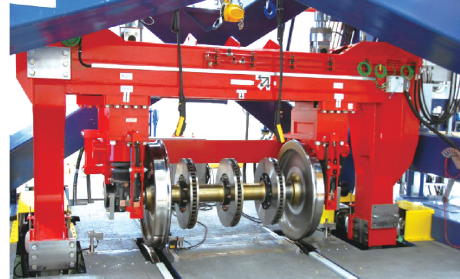


图 3 高速轮对滚振试验台试件安装图

Fig. 3 Specimen installation diagram of high speed wheelset rolling vibration test bench

试验首先通过电火花加工对动车组轴箱的 SKF 圆柱滚子轴承施加人工故障,故障分类施加在内圈外圈和滚动体上,并压装在右侧轴箱,左侧轴箱则使用正常轴承。根据故障施加的位置,并考虑在工程实际中可能的复合故障情形,试验共设置了 8 个组别:内圈故障,外圈故障、滚动体故障、内圈+外圈复合故障、内圈+滚动体复合故障、外圈+滚动体复合故障、内圈+外圈+滚动体复合故障和无故障正常轴承。在轮对平稳运行后采集数据,每次数据采集时间为一分钟,轮对运行速度级分别为 100 km/h、150 km/h 和 200 km/h。

如图 4 所示,试验在左、右轴箱的下部垂直安装了两个压电式振动加速度传感器,分别连接至如图 5 所示的振动数据采集器和上位机软件中。振动加速度传感器的采样频率为 12 kHz。



图 4 轴箱部位传感器布置

Fig. 4 Sensor layout around axle box

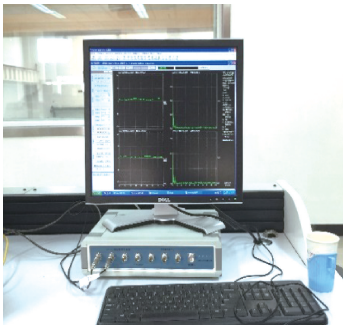


图5 振动数据采集仪

Fig. 5 Vibration data acquisition instrument

在动车组的实际运行中,无故障轴承的数量远远大于故障轴承的数量,出现复合故障的概率也要远远小于出现单一故障的概率。为模拟动车组轴箱轴承的实际故障数据量,以无故障轴承、单一故障、两种故障的复合故障、三种故障的复合故障=12:4:2:1的比例划分非平衡数据集^[14]。非平衡数据集的划分情况如表3所示。

表3 非平衡数据集划分情况

Table 3 Division of the unbalanced dataset

故障类型	故障编号	样本长度	样本数量(个)-工况/(km·h ⁻¹)		
			100	150	200
内圈故障	IF	400	400	400	400
外圈故障	OF	400	400	400	400
滚柱故障	RF	400	400	400	400
内圈+外圈故障	IOF	400	200	200	200
内圈+滚柱故障	IRF	400	200	200	200
外圈+滚柱故障	ORF	400	200	200	200
内圈+外圈+滚柱故障	IORF	400	100	100	100
无故障	N	400	1 200	1 200	1 200
总计			3 100	3 100	3 100

3 基于 shapelets 的动车组轴承故障诊断验证

模型参数对故障诊断的影响主要体现在 shapelets 的最小长度 L_{\min} 和超平面参数的个数 K (K 同时也等于所提取的 shapelets 的个数),其中 shapelets 的最小长度 L_{\min} 主要影响了子信号作为候选 shapelets 是否会因为长度过小或过大而影响计算的精度;超平面参数的个数决定了提取 shapelets 的个数,影响了诊断耗时。若初始化一个较大的 K 值,若 K 值取的过大,虽然在模型的后续更新过程中,多余的超平面参数会被更新到 0 附近,但会影响模型的收敛速度;若 K 的取值过小,则可能找不到一个合适的超平面对降维后的数组进行分类。如图 6 所示,若共有两个 shapelets,分别计算这两个 shapelets 与每条时间序列之间的相似度,则无法在数轴上找到一个阈值实现分类;而若映射到二维坐标系上,则可找到一条直线实现分类。

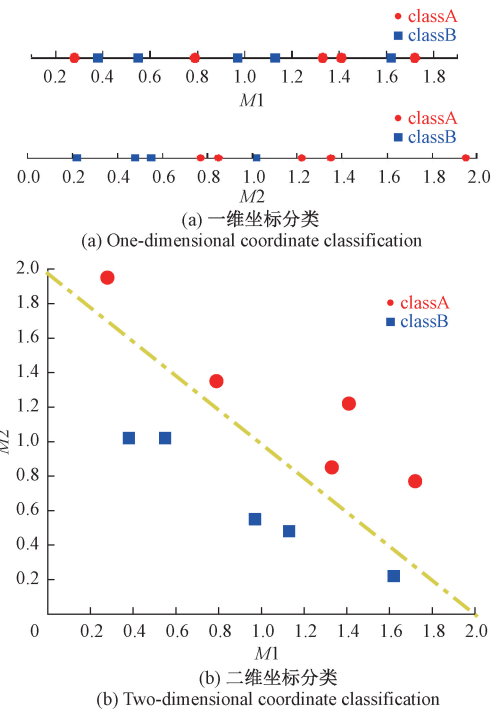


图6 超平面维度数目对分类结果的影响

Fig. 6 The influence of the number of hyperplane dimensions on the classification results

按照式(8)来初始化参数:

$$\begin{cases} Q = x_{\text{train}} \cdot \text{shape}^{[1]} \\ K = 24 \\ L_{\min} = \text{int}(0.2 \cdot Q) \\ \eta = 0.01 \\ \alpha = -100 \\ \text{epoch} = 15 \\ N = 2000 \end{cases} \quad (8)$$

取不同的 L_{\min} 和 K 值实现故障诊断,损失函数和诊断精度随迭代周期数的变化趋势分别如图 7 和图 8 所示。

综合上述诊断结果可以得出结论, L_{\min} 对诊断精度和模型收敛速度的影响是有限的,这说明 L_{\min} 的值过大或过小导致的 shapelets 模糊化提取不会对诊断的结果产生太大的误差。对于 K 的取值而言,当 K 值过小时,诊断精度完全无法满足工程实际的需要;当 $K \geq 15$ 时,模型的诊断性能有了较为明显的提升,诊断精度均达到了 95% 以上;随着 K 值的进一步增大,模型的收敛速度加快,在 $K = 24$ 时的损失函数和诊断精度的变化曲线已经非常平滑;当 $K = 30$ 时,诊断精度的变化曲线发生轻微震荡,预计原因是 K 值过大而导致多余的超平面参数在收敛至 0 的过程中在 0 值的周围未能快速收敛。

当 L_{\min} 取 $0.2Q$, K 取 24 时,诊断模型提取出的 24 个 shapelets 如图 9 所示。

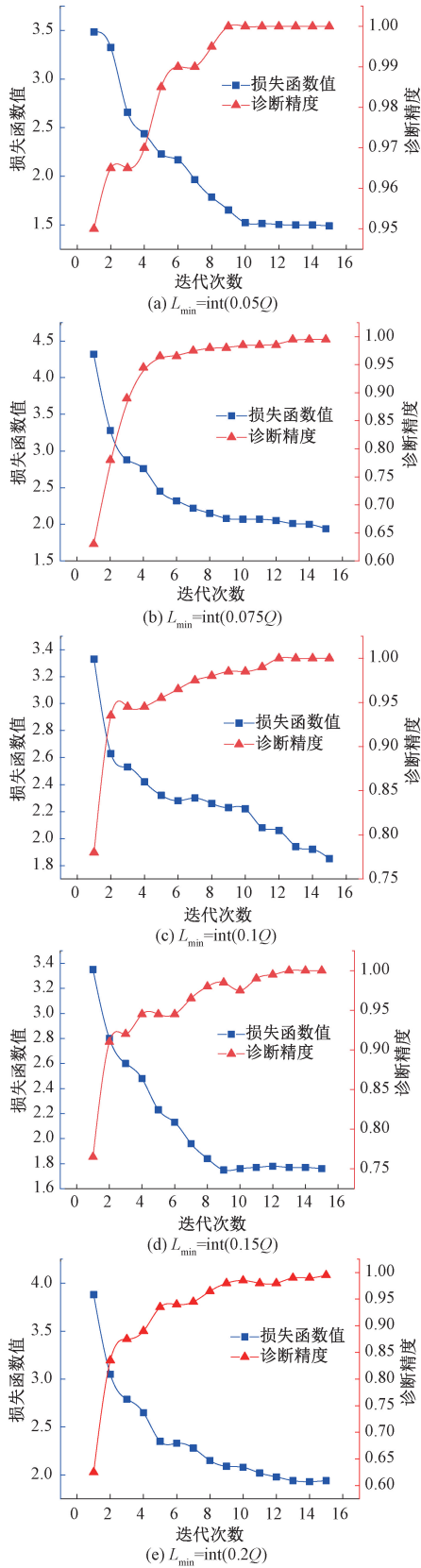


图 7 不同 L_{\min} 取值下学习模型的诊断精度变化趋势
 Fig. 7 Changing trend of the diagnosis accuracy of learning model under different L_{\min} values

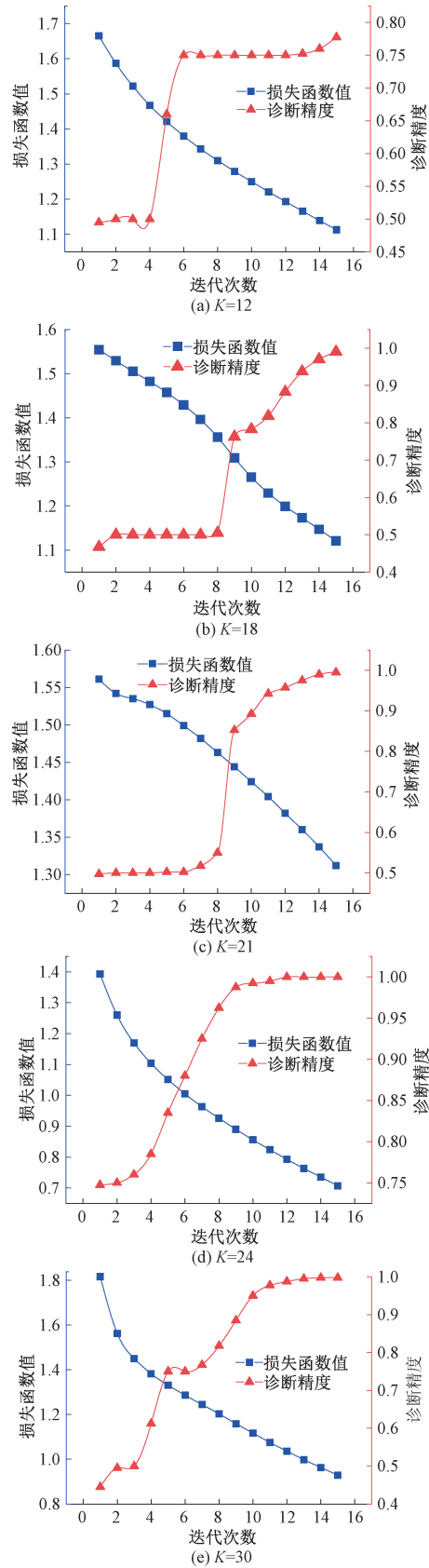


图 8 不同 K 取值下模型的诊断性能
 Fig. 8 Diagnostic performance of the model under different K values

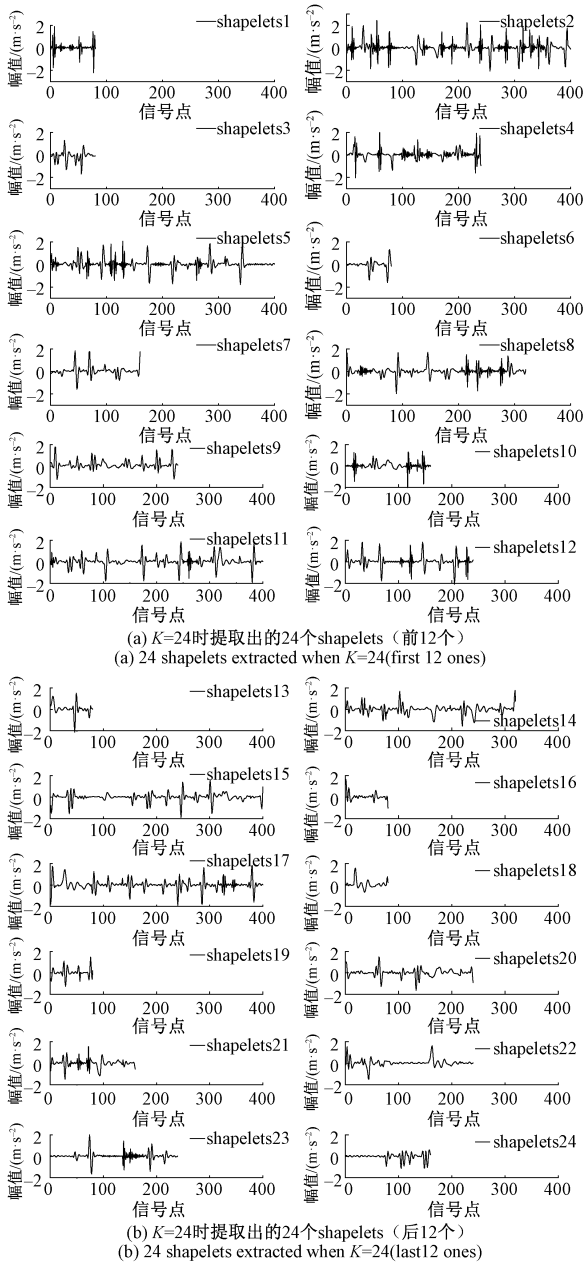


图9 K=24时提取出的24个shapelets
Fig. 9 24 shapelets extracted when K=24

4 提升模型泛化性能的方式

过拟合是机器学习中最常见的问题,具体到故障诊断领域,即模型能够在训练集上获得很高的诊断精度,但在测试集或其他新鲜样本上的诊断精度不佳。过拟合问题在滚动轴承故障诊断中的体现在于,以某一特定工况下的故障数据(如同转速,同温度,同轴承类型)为训练集的故障模型,可以有效对同工况下的故障实现故障诊断,但在处理测试集上的新鲜数据或不同工况下的轴承故障时,会产生诊断精度下降的情况。本文以 Dropout 思想作为启发,对模型进行改进。

Dropout 是一种常见的机器学习模型改进方法^[15],通过在训练神经网络时将一定比例的神经元输出置零,人为地阻止其参与反向传播。该方法可以减少算法对某些局部特征的依赖,重点学习模型中较为重要的特征,降低模型过高的学习能力,提升模型的泛化性能。Dropout 的示意图如图 10 所示。

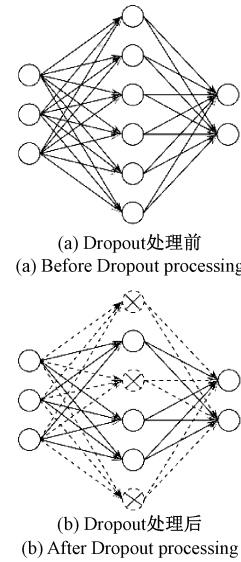


图 10 Dropout 处理方法

Fig. 10 Dropout processing method

由于本文中唯一的隐藏层 (Soft-min 层) 的输出具有明确的物理意义,即子信号与原信号之间的相似度,因此可以直接根据 Soft-min 层输出值的大小进行排序筛选,保留对时间序列相似度计算贡献较大的隐藏层输出,人为地将时间序列相似度较小的隐藏层置零,阻止其参与到反向传播算法中,本文设置 Dropout = 0.5,即将信号相似度较小的 50% Soft-min 输出置零。改进后的模型如图 11 所示。

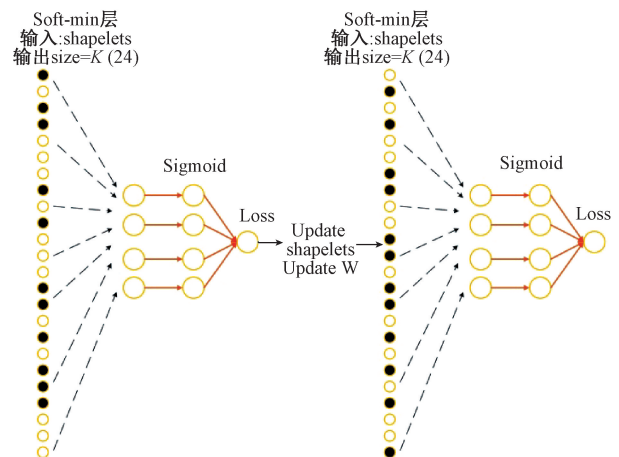


图 11 基于 Dropout 思想的改进模型

Fig. 11 Improved model based on Dropout idea

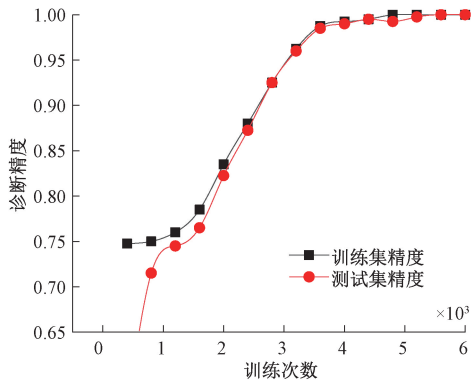


图 12 Dropout 处理后模型在训练集和测试集上诊断结果

Fig. 12 The diagnosis results of the model on the training set and test set after dropout processing

Dropout 处理之后的模型在训练集上和测试集上的表现如图 12 所示。可以发现模型在训练集和测试集上的诊断精度都达到了 100%,且在训练过程中没有出现训练集精度远远高于测试集精度的情况。

5 结 论

针对基于数据驱动的故障诊断方法存在的可解释性差,提取特征不具备物理特征的问题,本文将基于 shapelets 学习算法的时间序列分类方法引入故障诊断领域,通过动车组轮对台架滚振试验建立动车组轴箱轴承故障数据集,在该数据集有针对性地诊断模型的待定参数进行了确定,并实现了轴箱轴承的高精度诊断;针对机器学习方法中时常存在的过拟合问题,本文从 Dropout 理论中获得启发,对模型进行了类似 Dropout 的优化,优化后的诊断模型在训练集和测试集上均能去的较高的诊断精度。本文的研究结论如下:

1) 基于 shapelets 学习算法的时间序列分类方法是一种可行的轴承故障诊断方法,诊断精度高,模型的综合性能强大。将轴承的故障诊断问题转化为故障信号的相似度度量 and 分类问题,体现了 shapelets 作为“最具有代表性的时间序列子序列”的强可解释性。

2) 针对现存的故障诊断新方法对 CWRU 等公开数据的依赖性较大的不足,本文基于动车组轮对台架滚振实验,建立了动车组轴箱轴承故障数据集,并在该数据集上有针对性地 shapelets 学习算法中的待定参数进行了确定,最终实现了动车组轴箱轴承故障的高精度诊断。

3) 针对机器学习算法中的过拟合问题,本文收 Dropout 思想启发,人工地将隐藏层中对时间序列相似度计算贡献较小的神经元输出人为置零,减少模型对非重要故障信号数据特征的过强学习性能。改进后的模型在同一工况下的测试集和训练集上均达到了 100% 的诊断精度。

参考文献

- [1] 古莹奎,曾磊,张敏,等. 基于 CEEMDAN-SQI-SVD 的齿轮箱局部故障特征提取[J]. 仪器仪表学报, 2019, 40(5): 78-88.
GU Y K, ZENG L, ZHANG M, et al. Feature extraction method for gearbox local fault based on CEEMDAN-SQI-SVD [J]. Chinese Journal of Scientific Instrument, 2019, 40(5): 78-88.
- [2] 辛玉,李舜酩,王金瑞,等. 基于迭代经验小波变换的齿轮故障诊断方法[J]. 仪器仪表学报, 2018, 39(11): 79-86.
XIN Y, LI SH M, WANG J R, et al. Gear fault diagnosis method based on iterative empirical wavelet transform [J]. Chinese Journal of Scientific Instrument, 2018, 39(11): 79-86.
- [3] 刘长良,武英杰,甄成刚. 基于变分模态分解和模糊 C 均值聚类的滚动轴承故障诊断[J]. 中国电机工程学报, 2015, 35(13): 3358-3365.
LIU CH L, WU Y J, ZHEN CH G. Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C means clustering [J]. Proceedings of the CSEE, 2015, 35(13): 3358-3365.
- [4] 陈超,沈飞,严如强. 改进 LSSVM 迁移学习方法的轴承故障诊断[J]. 仪器仪表学报, 2017, 38(1): 33-40.
CHEN CH, SHEN F, YAN R Q. Enhanced least squares support vector machine-based transfer learning strategy for bearing fault diagnosis [J]. Chinese Journal of Scientific Instrument, 2017, 38(1): 33-40.
- [5] 曲建岭,余路,袁涛,等. 基于一维卷积神经网络的滚动轴承自适应故障诊断算法[J]. 仪器仪表学报, 2018, 39(7): 134-143.
QU J L, YU L, YUAN T, et al. Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network [J]. Chinese Journal of Scientific Instrument, 2018, 39(7): 134-143.
- [6] FENG J, LEI Y G. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization [J]. Mechanical Systems and Signal Processing, 2018(110): 349-367.
- [7] YE L, KEOGH E J. Time series shapelets: A new primitive for data mining [C]. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009: 947-956.

- [8] 王志海,张伟,原继东,等. 一种基于 Shapelets 的懒惰式时间序列分类算法[J]. 计算机学报,2019,42(1):29-43.
WANG ZH H, ZHANG W, YUAN J D, et al. A novel lazy time series classification algorithm based on the shapelets[J]. Chinese Journal of Computers, 2019, 42(1):29-43.
- [9] 原继东,王志海,韩萌,等. 基于逻辑 shapelets 转换的时间序列分类算法[J]. 计算机学报,2015,38(7):1448-1459.
YUAN J D, WANG ZH H, HAN M, et al. A logical shapelets transformation for time series classification[J]. Chinese Journal of Computers, 2015,38(7):1448-1459.
- [10] 杨艳林,叶枫,吕鑫,等. 一种基于 DTW 聚类的水文时间序列相似性挖掘方法[J]. 计算机科学,2016,43(2):245-249.
YANG Y L, YE F, LYU X, et al. DTW clustering-based similarity mining method for hydrological time series[J]. Chinese Journal of Computers, 2016,43(2):245-249.
- [11] 朱天,白似雪. 基于模式距离度量的时间序列相似性搜索[J]. 微计算机信息,2007(30):216-217.
ZHU T, BAI S X. Similarity search based on pattern distance measure in time-series database [J]. Microcomputer Information, 2007(30):216-217.
- [12] 王少鹏,闻英友,赵宏. 基于 LCSS 的数据流相似性查询处理算法研究[J]. 计算机研究与发展,2015,52(9):1976-1991.
WANG SH P, WEN Y Y, ZHAO H. Similarity query processing algorithm over data stream based on LCSS[J]. Journal of Computer Research and Development, 2015,52(9):1976-1991.
- [13] GRABOCKA J, SCHILLING N, WISTUBA M, et al. Learning time-series shapelets[M]. ACM,2014.
- [14] 荆云建. 基于改进型卷积神经网络的动车组滚动轴承故障诊断方法研究[D]. 北京:北京交通大学,2019.

JIN Y J. Research on fault diagnosis method for rolling bearing of EMU based on improved convolutional neural network[D]. Beijing:Beijing Jiaotong University,2019.

- [15] HINTON G E, SRIVASTAVA N, KRIZHEVAKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Ence, 2012, 3(4):212-223.

作者简介



宋志坤(通信作者),1995年于东北石油大学获得学士学位,1998年于北京交通大学获得硕士学位,2011年于北京交通大学获得博士学位,现为北京交通大学机电学院副教授,主要研究方向为车辆系统动力学和车辆零部件疲劳、断裂、故障诊断等。

E-mail:zhksong@bjtu.edu.cn

Song Zhikun (Corresponding author) obtained bachelor degree from Northeast Petroleum University in 1995, and received master degree in 1998 and Ph. D. degree in 2011 both from Beijing Jiaotong University. He is currently an associate professor at School of Mechanical, Electrical and Control Engineering, Beijing Jiaotong University. His main research interest includes vehicle system dynamics, vehicle part fatigue, fracture and fault diagnosis, etc.



徐立成,2018年于北京交通大学获得学士学位,2020年于北京交通大学获得硕士学位,现为中车戚墅堰机车车辆工艺研究所有限公司助理工程师,主要研究方向为智能机械传动技术。

E-mail:LichengXu616@163.com。

Xu Licheng obtained bachelor degree in 2018 and master degree in 2020 both from Beijing Jiaotong University. He is currently an assistant engineer at CRRC Qishuyan Institute Co., Ltd. His main research field is intelligent mechanical transmission technology.