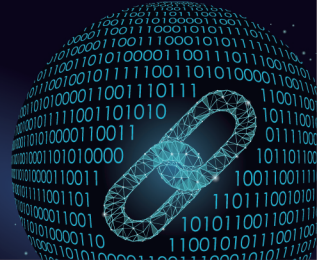




中国科技期刊卓越行动计划项目入选期刊

# 控制与决策

CONTROL AND DECISION



## 基于深度残差概率随机森林的时间序列分类方法

刘颖, 李旭, 吕政, 赵琨, 王伟

引用本文:

刘颖,李旭,吕政,赵,王伟. 基于深度残差概率随机森林的时间序列分类方法[J]. *控制与决策*, 2024, 39(7): 2315–2324.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1752>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于生成对抗网络的大规模路网交通流预测算法

Traffic flow forecasting algorithm for large-scale road network based on GAN

控制与决策. 2021, 36(12): 2937–2945 <https://doi.org/10.13195/j.kzyjc.2020.0333>

#### 基于FWADE-ELM的短时交通流预测方法

Short-term traffic flow forecasting based on hybrid FWADE-ELM

控制与决策. 2021, 36(4): 925–932 <https://doi.org/10.13195/j.kzyjc.2019.1103>

#### 一种基于深度学习的时间序列预测方法

A time series prediction method based on deep learning

控制与决策. 2021, 36(3): 645–652 <https://doi.org/10.13195/j.kzyjc.2019.0809>

#### 基于改进卷积神经网络的动力下肢假肢运动意图识别

Intent recognition of power lower-limb prosthesis based on improved convolutional neural network

控制与决策. 2021, 36(12): 3031–3038 <https://doi.org/10.13195/j.kzyjc.2020.0326>

#### 基于FWADE-ELM的短时交通流预测方法

Short-term traffic flow forecasting based on hybrid FWADE-ELM

控制与决策. 2021, 36(4): 925–932 <https://doi.org/10.13195/j.kzyjc.2019.1103>

# 基于深度残差概率随机森林的时间序列分类方法

刘颖, 李旭, 吕政<sup>†</sup>, 赵珺, 王伟

(大连理工大学控制科学与工程学院, 辽宁大连 116024)

**摘要:** 时间序列数据广泛存在于工业、医疗等应用领域, 由于其时序相关性强、特征空间维度大, 使得传统的时间序列分类方法普遍存在精度不足和需要复杂特征工程等问题. 充分考虑深度神经网络在处理复杂时序数据上的优越性以及决策树方法拟合数据能力强的优势, 提出一种基于残差网络和概率决策树的端到端统一深度学习模型. 该模型利用残差网络从原始时间序列中提取高级特征, 为了更好地建立时序数据特征与类别标签间的映射关系, 将概率决策树融入至残差网络的分类层. 同时, 设计随机子空间的集成策略, 缓解由于残差网络的深层结构产生的过度拟合现象, 并给出联合优化模型分裂参数和预测参数的迭代优化方案. 在大量的基准数据集和工业案例上进行实验和分析, 实验结果表明, 所提出模型的分类性能优于传统方法与其他深度学习方法, 且可有效提高残差网络的泛化能力.

**关键词:** 时间序列分类; 端到端学习; 残差网络; 概率决策树; 随机子空间

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1752

**引用格式:** 刘颖, 李旭, 吕政, 等. 基于深度残差概率随机森林的时间序列分类方法[J]. 控制与决策, 2024, 39(7): 2315-2324.

## Time series classification using deep residual probability random forests

LIU Ying, LI Xu, LV Zheng<sup>†</sup>, ZHAO Jun, WANG Wei

(School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China)

**Abstract:** Time series data widely exists in industrial, medical and other application fields. Due to its strong temporal correlation and large feature space dimension, traditional time series classification methods generally have problems of insufficient accuracy and complex feature engineering. This paper proposes an end-to-end unified deep learning model based on residual networks and probability decision trees by fully considering the superiority of deep neural networks in dealing with complex time series data and the strong ability of a decision tree method to fit data. This model uses a residual network to extract advanced features from original time series. In order to better establish the mapping relationship between features and labels, probability decision trees are integrated into the classification layer of the residual network. Meanwhile, the integration strategy of random subspace is designed to alleviate the over-fitting phenomenon caused by the deep structure of the residual network. We also give the iterative optimization scheme to jointly optimize model's split parameters and prediction parameters. Experiments and analysis on a large number of benchmark datasets and an industrial case show that the classification performance of the proposed model is better than that of traditional methods and other deep learning methods, and the generalization ability of the residual network can be improved effectively.

**Keywords:** time series classification; end-to-end learning; residual network; probability decision tree; random subspace

## 0 引言

物联网的新时代背景下, 时间序列数据广泛存在于金融学、医学以及工业界等应用领域. 时间序列数据具有特征空间维度高、数据依赖关系复杂等特点, 因此, 对时序数据的分类是一项十分具有挑战性的问题. 目前, 研究者们提出了多种时间序列分类算法, 包括基于距离的方法、基于特征的方法、基于集成学习

的方法等. 早期的基于距离的方法本质是采用欧氏距离、动态时间规整<sup>[1]</sup>等对时序样本进行相似性度量, 并结合分类器实现分类. 考虑到现有的时间序列相似性度量方法复杂度高的问题, Li等<sup>[2]</sup>通过矩阵概要文件(matrix profile, MP)算法构造了子序列距离矩阵, 从而能够快速找到不同时间序列间最相似的子序列. 基于特征的方法核心在于如何有效地提取原始

收稿日期: 2022-10-08; 录用日期: 2023-04-03.

基金项目: 国家自然科学基金项目(61873048, 62003072); 国家科技部重点研发计划项目(2017YFA0700300); 中央高校基本科研业务费专项资金项目(DUT22JC16).

<sup>†</sup>通讯作者. E-mail: lvzheng@dlut.edu.cn.

时序数据中的辨别性特征,目前已有许多特征提取方法,如时间序列判别性子序列的提取<sup>[3]</sup>、时间序列特征的符号化<sup>[4-5]</sup>等.文献[6]通过将时间序列转化为复杂网络,利用复杂网络的拓扑特征提取能够更好地代表原始时间序列的动态特征.尽管相关文献验证了该类方法能够基于提取的全局和局部信息提高分类精度,但是,复杂且繁重的特征提取工程仍然是不可避免的.文献[7-8]采用集成学习策略组合不同的时序特征和分类器以实现更高的分类精度,然而,基于集成的方法分类精度受基本分类器的影响,且极高的时间复杂度使其难以部署于实时应用领域.

由于传统方法均伴随繁杂的特征提取和数据预处理的过程,越来越多的研究人员利用深度学习技术实现端到端的时间序列分类.Wang等<sup>[9]</sup>提出了3个端到端基线模型,利用MLP、FCN和ResNet对时序样本进行分类.Karim等<sup>[10]</sup>充分结合LSTM能够获取历史数据蕴含信息的特性和FCN擅于提取时序局部特征的优势,建立了一个端到端混合模型LSTM-FCN.Ismail等<sup>[11]</sup>综合评价了时间序列分类的深度学习方法,研究表明,CNN因其能够通过时序卷积充分提取各种特征而得到了广泛应用.鉴于ESN优越的时序建模能力,文献[12]提出了一种用于时间序列分类的端到端回声记忆增强网络模型(echo memory-augmented network, EMAN),通过回声记忆增强机制成功解决了ESN难以捕获时序数据长期依赖性的问题.随着特征融合策略表现出强大的时序分类能力,Wang等<sup>[13]</sup>提出了自适应特征融合网络(adaptive feature fusion network, AFFNet)来提高时序分类的准确性,该模型充分融合了多尺度时间特征和距离特征.尽管深度学习方法克服了传统方法繁杂特征工程的弊端,但是,复杂的结构和强大的特征提取能力使得模型易在小规模时间序列数据集上过拟合,因此,提高泛化能力是深度学习方法所面临的关键问题<sup>[14]</sup>.

决策树是数据挖掘领域中常用的分类工具.传统决策树在分裂节点处通过对特征空间的“硬”划分

寻找最优分裂边界,将数据样本从根节点递归地划分到叶节点,但是,这种启发式贪婪算法易陷入局部最优的困境,且“硬”划分方式使得决策树难以融入深度学习框架中实现端到端的学习模式.对此,Kontschieder等<sup>[15]</sup>提出了深度神经决策林(deep neural decision forests, DNDFs),通过在分裂节点处引入概率分裂函数,保证了决策树的分裂函数对于特征向量具有可微性,从而允许分裂节点与深度神经网络联合学习,并在图像识别等各种视觉任务上取得了显著的优势.尽管时序问题与视觉问题具有学习空间特征信息的相似性,但是,鲜有将深度学习模型在特征提取方面的优势与随机森林浅层模型的良好分类性能相结合来解决时间序列分类问题的研究.

鉴于以上传统时间序列分类方法存在的问题以及基于深度学习的端到端方法的优越性,本文提出一种基于残差网络和概率随机森林的端到端时间序列分类模型DRPRFs(deep residual probability random forests).该方法通过ResNet挖掘时间序列的深层特征信息,并利用集成模型概率随机森林替换网络的softmax分类器,通过对提取的高级非线性特征随机划分特征子空间来达到缓解模型过拟合现象的目的.分类实验结果表明,与传统方法相比,所提出方法能够在避免繁琐的数据预处理和特征加工的情况下获得较好的分类性能,同时,能够有效地提高深度神经网络在时间序列分类问题上的泛化能力.

## 1 背景知识

### 1.1 残差网络

残差网络(residual network, ResNet)通过引入恒等映射的设计<sup>[16]</sup>,成功缓解了网络由于深度增加带来的梯度爆炸或梯度消失的问题,在图像分类、语音识别等领域获得了广泛应用.Wang等<sup>[9]</sup>首次探索了ResNet在时间序列分类任务上的性能,其结构如图1所示,通过在卷积块的输入与输出间添加残差连接,将全卷积网络(fully convolutional networks, FCN)扩展为具有3个残差块的深层网络结构.

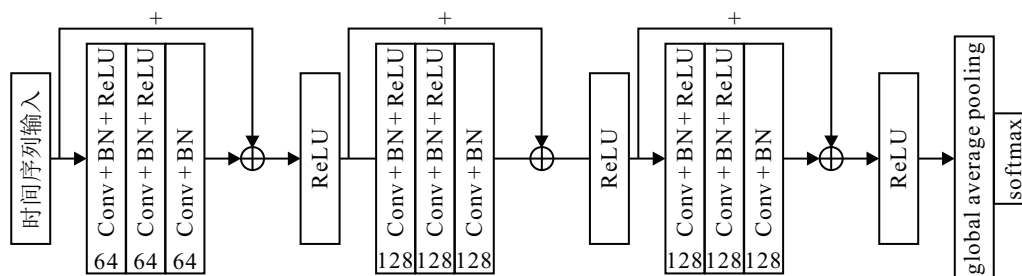


图1 残差网络

FCN使用卷积神经网络作为特征提取模块,一个基本的卷积块包括卷积层、批量归一化层和ReLU激活函数. 其中:卷积层负责提取数据的局部特征,卷积后通过批量归一化层和激活函数实现非线性变换,有

$$\begin{cases} t = W \otimes x + b, \\ a = \text{BN}(t), \\ y = \text{ReLU}(a). \end{cases} \quad (1)$$

式中: $\otimes$ 为卷积运算符; $W$ 和 $b$ 为卷积核的权值和偏置,堆叠3个卷积块构建一个完整的全卷积网络.当输入为单变量时间序列(univariate time series, UTS)时,FCN选用3个大小依次为 $\{8, 5, 3\}$ 的一维卷积核;当输入为多变量时间序列(multivariate time series, MTS)时,卷积核的维数选择与MTS的维数相同.在此基础上,ResNet通过在FCN的输入与输出间添加残差连接,堆叠3个FCN构成ResNet.一个基本的残差块可表示为

$$\begin{cases} h_1 = \text{Block}_{k_1}(x), \\ h_2 = \text{Block}_{k_2}(h_1), \\ h_3 = \text{Block}_{k_3}(h_2), \\ h = \text{ReLU}(h_3 + x). \end{cases} \quad (2)$$

其中3个残差块的滤波器数量分别设置为64、128和128.最终,将经过逐层映射的高级特征输入至全局平均池化层,再由softmax层进行时序分类.

### 1.2 概率决策树

概率决策树通过在分裂节点处引入概率分裂函数允许样本以一定的概率路由到左右子节点.图2是一个深度为3的概率决策树.

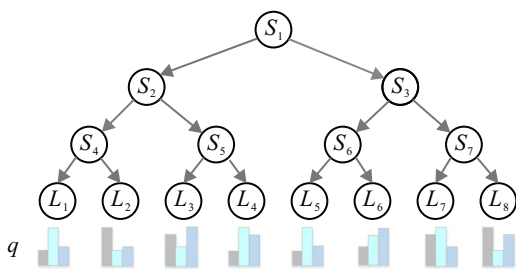


图2 概率决策树

给定输入空间 $X \subset \mathbb{R}^p$ 有 $p$ 维特征,输出空间 $Y = \{1, 2, \dots, C\}$ .假设一棵概率决策树由一组分裂节点 $N$ 和一组叶子节点 $L$ 组成,每个分裂节点 $n \in \{1, 2, \dots, N\}$ 采用sigmoid函数执行概率分裂,样本 $x$ 路由到左子节点的概率为

$$s_n(x, \theta_n) = \sigma(f_n(x, \theta_n)), \quad (3)$$

其中 $f_n(x, \theta_n) = (x^T, 1) \cdot \theta_n$ 为依赖于样本 $x$ 的 $p$ 维特征向量和分裂参数 $\theta_n$ 的实值特征学习函数.在概率决策树中,每个样本可通过独特的路径到达每个叶节点,定义样本 $x$ 从根节点路由到叶节点 $l \in \{1, 2, \dots, L\}$ 的概率为

$$p(l|x; \Theta) = \prod_{n \in N} s_n(x; \theta_n)^{I(l \in L_n^l)} \cdot (1 - s_n(x; \theta_n))^{I(l \in L_n^r)}. \quad (4)$$

式中: $\Theta = (\theta_1, \theta_2, \dots, \theta_N)$ 为分裂参数集, $I(\cdot)$ 为指示函数, $L_n^l$ 和 $L_n^r$ 分别为节点 $n$ 的左子树和右子树所拥有的叶节点的集合.概率决策树的每个叶节点 $l$ 均存储输出空间 $Y$ 的类分布 $q_l = (q_{l1}, q_{l2}, \dots, q_{lC})$ .其中: $q_{lc} \in [0, 1]$ 且 $\sum_{c=1}^C q_{lc} = 1$ ,这样概率决策树的预测结果由路由概率和对应的叶节点的类分布相乘得到

$$g(x; q, \Theta) = \sum_{l \in L} p(l|x; \Theta) \cdot q_l. \quad (5)$$

## 2 DRPRFs模型的构建

时间序列是在相等时间间隔内对观测对象的一系列有序实值数据,根据观测对象数量分为UTS和MTS.假设给定一个样本量大小为 $I$ 的时间序列数据集 $D = \{X_1, X_2, \dots, X_I\}$ ,每个时间序列表示为 $X = (x^1, x^2, \dots, x^u)^T$ 且 $x^w = (x_1^w, x_2^w, \dots, x_v^w)$ .其中: $u$ 为变量维度, $v$ 为时间维度, $x^w$ 为观测对象 $w$ 的 $v$ 个观测值.每个时间序列样本 $X_i$ 对应于类标 $Y_i \in \{1, 2, \dots, C\}$ , $C$ 为类别数目.

### 2.1 模型描述

DRPRFs模型如图3所示,主体为残差网络和概率随机森林的端到端统一深度架构. ResNet模块负

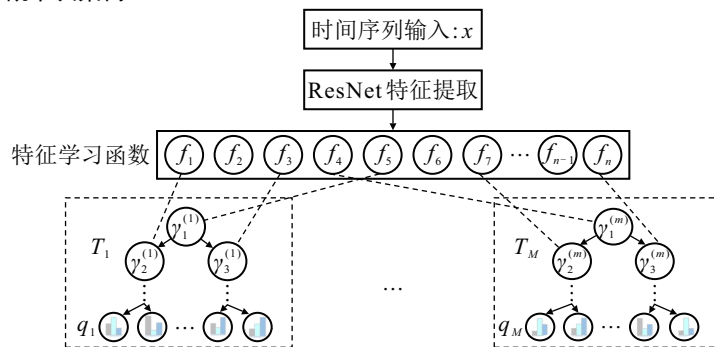


图3 深度残差概率随机森林

责从复杂的原始时间序列中提取非线性判别特征, 概率随机森林模块根据这些深层特征信息生成样本的概率预测结果. 本文定义了一个基于森林的全局损失函数, 通过联合优化所有树的分裂参数和预测参数来最小化损失函数, 从而得到模型的全局最优解.

### 2.1.1 深度残差概率随机森林

假设 DRPRFs 由  $M$  棵概率决策树  $F = \{T_1, T_2, \dots, T_M\}$  组成, 时序样本  $X$  经 ResNet 的时序卷积和池化等操作映射至高级特征, 令  $f(X, \Theta)$  为输入样本与网络提取的特征间的映射函数, 则  $f_n(X, \Theta)$  为全连接层的第  $n$  个特征学习函数.

通过由不同的特征组成随机子空间训练集成模型中的基分类器, 降低每个基分类器的相关性, 从而提高模型的泛化能力. 本文将网络提取的特征集随机划分为  $M$  个特征子空间  $f^{(m)} (m = 1, 2, \dots, M)$ , 并引入索引函数  $\gamma^{(m)} (m = 1, 2, \dots, M)$  构建特征子空间与决策树的映射关系, 如下式所示:

$$\gamma^{(m)} = f^{(m)}, \quad (6)$$

从而保证特征学习函数随机输入至决策树的分裂节点, 决策树  $m$  的分裂节点  $n$  概率分裂函数为

$$s_n^{(m)} = \sigma(f_{\gamma_n^{(m)}}(X, \theta_n^m)). \quad (7)$$

DRPRFs 的目标在于优化整个模型的损失函数, 获取全局最优参数, 进而对测试样本进行概率预测. 不同于 DNDFs 定义单棵决策树的损失函数并以串行的方式优化决策林中的每棵树的学习策略, DRPRFs 明确定义了基于整个森林的全局损失函数. 对于概率决策树  $m$ , 定义  $\Theta^m = (\theta_1^m, \theta_2^m, \dots, \theta_N^m)$  为分裂参数集,  $q^m = (q_1^m, q_2^m, \dots, q_L^m)$  为预测参数集, 选择训练数据的对数损失

$$\begin{aligned} R_m(q^m, \Theta^m) = & - \sum_{i=1}^I \log[g_m(X_i; q^m, \Theta^m)] = \\ & - \sum_{i=1}^I \sum_{l=1}^L \log[p(l|X_i; \Theta^m) \cdot (q_l^m)_{Y_i}] \end{aligned} \quad (8)$$

为损失函数, 因此, DRPRFs 的全局损失函数为

$$R_F(q, \Theta) = \frac{1}{M} \sum_{m=1}^M R_m(q^m, \Theta^m). \quad (9)$$

通过最小化损失函数, 得到模型的最优参数  $(q^*, \Theta^*)$ , 即

$$(q^*, \Theta^*) = \arg \min_{q, \Theta} R_F(q, \Theta). \quad (10)$$

在测试阶段, 时间序列样本会通过特定的路径路由到每棵树的所有叶节点, 每棵树会为样本提供一个

预测的标签分布, 最终将所有树预测结果的平均值作为模型的输出, 有

$$g(X; q, \Theta, F) = \frac{1}{M} \sum_{m=1}^M g_m(X; q^m, \Theta^m), \quad (11)$$

并选择模型输出的标签分布向量中描述值最大的标签作为模型的分类结果, 即

$$Y = \max_c g_c(X; q, \Theta, F). \quad (12)$$

### 2.1.2 模型优化

为了能够同时优化所有决策树的分裂参数  $\Theta$  和预测参数  $q$ , 本文给出基于全局损失函数的模型优化策略. 由于全局损失函数可拆分为  $\Theta$  相关部分和  $q$  相关部分, 进而能够分别优化  $\Theta$  和  $q$ , 使得全局损失函数最小化, 可采用两步交替迭代优化方案, 即先固定叶节点的类分布  $q$  优化  $\Theta$ , 然后固定  $\Theta$  更新叶节点的类分布  $q$ . DRPRFs 中所有分裂节点参数共享 ResNet 网络参数, 因此, 本文利用反向传播机制联合优化所有的概率决策树, 即

$$\Theta^{t+1} = \Theta^t - \frac{\alpha}{M} \sum_{m=1}^M \frac{\partial R_m}{\partial \Theta^m}. \quad (13)$$

其中:  $\alpha$  为学习率; 损失函数相对于  $\Theta$  的梯度值可由

$$\frac{\partial R_m}{\partial \Theta^m} = \sum_{i=1}^I \sum_{n \in N_m} \frac{\partial R_m}{\partial f_{\gamma_n^{(m)}}(x_i, \theta_n^m)} \frac{\partial f_{\gamma_n^{(m)}}(x_i, \theta_n^m)}{\partial \theta_n^m} \quad (14)$$

计算得到,  $N_m$  和  $\gamma^{(m)}$  分别为决策树  $m$  的分裂节点集和对应的特征子空间, 然后通过梯度下降法来最小化关于  $\Theta$  的损失项. 由于每棵树的叶子节点的类分布是相对独立的, 可独立更新每棵树的预测参数. 当固定  $\Theta$  时, 对  $q$  的优化问题转变为解决凸优化问题, 即

$$\begin{aligned} \min_{q^m} R_m(q^m, \Theta^m); \\ \text{s.t. } \forall l \in L, \sum_{c=1}^C q_{lc} = 1. \end{aligned} \quad (15)$$

本文遵循 DNDFs 的思路利用变分边界法解决上述凸优化问题, 使得损失函数快速收敛. 基于交替迭代优化方案, 当 DRPRFs 模型达到最大迭代次数或收敛时, 得到模型的全局最优解, 从而实现对测试样本的最优决策. DRPRFs 算法的整体流程如图 4 所示.

## 2.2 时间复杂度

假设 DRPRFs 包含  $M$  棵深度为  $h$  的概率决策树, 即分裂节点数和预测节点数分别为  $2^h - 1$  和  $2^h$ , 给定样本数为  $I$  的时间序列. 由式 (14) 进行时间复杂度分析可知, DRPRFs 分裂节点的正向和反向传播的训练时间复杂度均为  $O((2^h - 1) \times C \times M \times I)$ , 其中  $C$  为类别数目. 对于预测节点而言, 需要求解所有类别的概

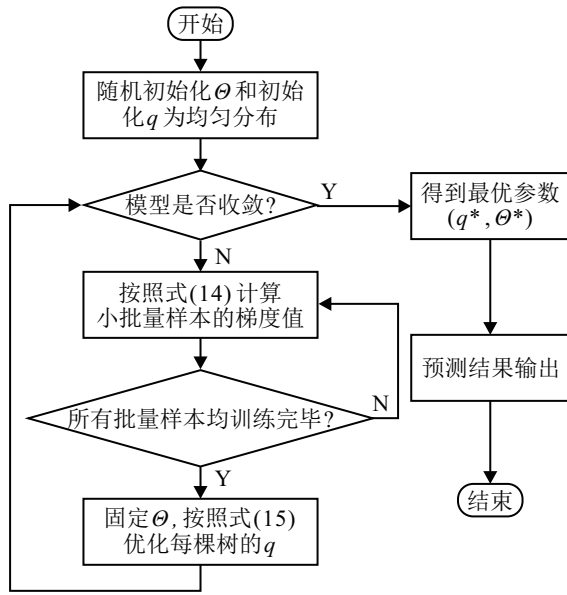


图4 DRPRFs算法流程

率分布,因此,迭代更新预测节点参数的时间复杂度为 $O(2^h \times C \times M \times I)$ . 综上所述,DRPRFs模型的整体时间复杂度为 $O(2^h \times C \times M \times I + (2^h - 1) \times C \times M \times I)$ ,简记为 $O(HCMI)$ ,其中 $H$ 表示 $2^h$ . 可以看出,所提出方法的时间复杂度与数据集规模、类别数、概率决策树的深度和数量密切相关.

### 3 实验与分析

为了验证所提出DRPRFs模型的分性能 and 泛化能力,本文通过在UCR单变量时间序列数据集<sup>[17]</sup>、UEA多变量时间序列数据集<sup>[18]</sup>和一个液压系统的故障诊断的工业数据集上与其他先进分类模型进行性能对比实验,以体现DRPRFs模型的优越性. 所有实验均是在Nvidia RTX 2080 Ti显卡、Intel(R) Xeon(R) E 5-2 680处理器、14 GB内存环境中执行的.

#### 3.1 标准数据集实验

##### 3.1.1 数据选取

标准数据集上的实验分为单变量时间序列分类和多变量时间序列分类两组实验,实验数据均来自UCR & UEA时间序列数据仓库. UTS分类实验选取UCR中45个不同领域的数据集进行实验,包括模拟、光谱仪、传感器、心电图、图像、运动以及设备这7种类型的数据,每个数据集默认地划分为训练集和测试集,具体信息如表1所示. 这些数据集具有变化多样的序列长度和形态特征,能够为所提出算法提供全面的评价. UCR档案对所有数据集均进行 $z$ 归一化处理,使得时间序列数据均值为0,方差为1,从而消除了量纲对实验的影响.

MTS分类实验采用UEA中20个等长序列的数据集进行实验,表2给出了这些数据集的基本信息.

表1 UTS数据集信息

数据集	训练数	测试数	序列长度	类别数	类型
ArrowHead	36	175	251	3	Image
BirdChicken	20	20	512	2	Image
CBF	30	900	128	3	Simulated
ClorineConc	467	3 840	166	3	Sensor
Coffee	28	28	286	2	Spectro
Computers	250	250	720	2	Device
DistPhalOutAgeGrp	400	139	80	6	Image
ECG 5 000	500	4 500	140	5	ECG
ECGFiveDays	23	861	136	2	ECG
FacesUCR	200	2 050	131	14	Image
Fish	175	175	463	7	Image
FordA	3 601	1 320	500	2	Sensor
GunPoint	50	150	150	2	Motion
Ham	109	105	431	2	Spectro
HandOutlines	1 000	370	2 709	2	Image
InlineSkate	100	550	1 882	7	Motion
ItalyPowerDemand	67	1 029	24	2	Sensor
LargeKitchenApp	375	375	720	3	Device
Lightning 2	60	61	637	2	Sensor
Meat	60	60	448	3	Spectro
MedicalImages	381	760	99	10	Image
MidPhalOutlineCor	600	291	80	2	Image
MoteStrain	20	1 252	84	2	Sensor
OSULeaf	200	242	427	6	Image
PhalangesOutCorrect	1 800	858	80	2	Image
Phoneme	214	1 896	1 024	39	Sensor
Plane	105	105	144	7	Sensor
ProxPhalOutAgeGrp	400	205	80	3	Image
RefrigerationDevices	375	375	720	3	Device
ScreenType	375	375	720	3	Device
SonyAIBORobSurf 1	20	601	70	2	Sensor
SonyAIBORobSurf 2	27	953	65	2	Sensor
StarLightCurves	1 000	8 236	1 024	3	Senesor
Strawberry	613	370	235	2	Spectro
SwedishLeaf	500	625	128	15	Image
SyntheticControl	300	300	60	6	Simulated
ToeSegmentation1	40	228	277	2	Motion
Trace	100	100	275	4	Sensor
TwoLeadECG	23	1 139	82	2	ECG
TwoPatterns	1 000	4 000	128	4	Simulated
UwaveGestLibZ	896	3 582	315	8	Motion
Wafer	1 000	6 164	152	2	Sensor
Worms	181	77	900	5	Motion
WormsTwoClass	181	77	900	2	Motion
Yoga	300	3 000	426	2	Image

数据集来源于人类活动识别、心电/脑电图分类等多个应用领域,数据类别的范围为2~26,多变量的维度包含在2~28维之间,时间序列长度从8~2 500不等.

##### 3.1.2 实验设置

基于UCR & UEA数据集分析对比所提出模型与目前多个文献公布的先进算法的分类效果,对比算法涉及基于距离、特征、集成学习和深度学习的多种TSC方法. UTS分类采用的对比算法包括:ED、DTW、BOSS<sup>[4]</sup>、WEASEL<sup>[19]</sup>、COTE<sup>[7]</sup>、LSTM-FCN<sup>[10]</sup>、EMAN<sup>[12]</sup>、AFFNet<sup>[13]</sup>和ResNet<sup>[9]</sup>; MTS分类采用的对比算法包括:ED、DTW<sub>D</sub><sup>[20]</sup>、WEASEL+MUSE<sup>[21]</sup>、

表2 MTS数据集信息

数据集	训练数	测试数	变量数	序列长度	类别数
ArticularyWordRecognition	270	300	9	144	25
BasicMotions	40	40	6	100	4
CharacterTrajectories	1 422	1 436	3	182	20
Cricket	108	72	6	1 197	12
Epilepsy	137	138	3	206	4
ERing	30	270	4	65	6
FingerMovements	316	100	28	50	2
HandMovementDirection	160	74	10	400	4
Handwriting	150	850	3	152	26
JapaneseVowels	270	370	12	29	9
Libras	180	180	2	45	15
LSST	2 459	2 466	6	36	14
NATOPS	180	180	24	51	6
PenDigits	7 494	3 498	2	8	10
Phoneme	3 315	3 353	11	217	39
RacketSports	151	152	6	30	4
SelfRegulationSCP 1	268	293	6	896	2
SelfRegulationSCP 2	200	180	7	1 152	2
StandWalkJump	12	15	4	2 500	3
UWaveGestureLibrary	120	320	3	315	8

MLSTM-FCN<sup>[22]</sup>、TapNet<sup>[23]</sup>、ShapeNet<sup>[24]</sup>和ResNet<sup>[9]</sup>.

在实验中,残差网络模块的参数详情参见第1.1节,本文设置特征学习函数的输出单元数为128.为了避免复杂的参数寻优过程,概率随机森林默认拥有5棵决策树,树的深度设为6.实验对模型进行1 500次迭代训练,使用Adam optimizer联合优化模型参数,初始学习率设为0.001,批尺寸大小设置为16.

### 3.1.3 分类性能评估

表3为UTS分类实验结果.对每个数据集的最佳分类精度进行黑体加粗显示,表3的底部统计比较了各模型的平均分类精度和胜出率,其中胜出率为该方法取得最佳分类精度的数据集个数与实验数据集总个数的比值.分析表3可知,DRPRFs的平均分类精度高于其他模型,且在UCR数据集上胜出的个数最多,表明所提出模型在UTS分类问题上的整体优越性.深度学习模型在时间序列分类问题上往往由于数据量不足而导致数据分析性能严重下降,过拟合更易发生,不能保证较高的分类精度.在ECGFiveDays、Fish、Lightning 2等几个训练集规模很小的数据集上,所提出模型表现出优于ResNet的分类效果,表明DRPRFs的“随机子空间”集成策略可有效地提高深度网络结构在小数据集上的泛化能力.MTS分类实验结果如表4所示.DRPRFs在UEA的20个数据集集中的9个数据集获得了最优精度,明显优于传统算法与深度学习算法.相较于UTS,MTS的时空特征更为丰富,深度学习算法能够更好地将特征空间映射至所属类标签,同时解决传统方法在解决多变量时间序列分类问题中的空间维度的困扰.对比MLSTM-FCN、TapNet与ShapeNet这些深度模型,DRPRFs仍然具有优越的分类性能.

表3 UTS分类实验结果对比

数据集	ED	DTW	BOSS	WEASEL	COTE	LSTM-FCN	EMAN	AFFNet	ResNet	DRPRFs
ArrowHead	0.800	0.703	0.834	<b>0.873</b>	0.811	0.829	0.846	0.800	0.817	0.771
BirdChicken	0.550	0.750	0.950	0.900	0.900	0.910	0.900	0.900	0.900	<b>1.000</b>
CBF	0.852	0.997	0.998	0.993	0.996	0.998	<b>1.000</b>	0.994	0.994	0.971
ClorineConc	0.650	0.648	0.661	0.728	0.727	0.803	0.796	<b>0.862</b>	0.828	0.847
Coffee	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
Computers	0.576	0.700	0.756	0.716	0.740	0.809	0.756	0.840	0.824	<b>0.860</b>
DistPhalOutAgeGrp	0.626	0.770	0.748	0.777	0.748	<b>0.868</b>	0.860	0.741	0.798	0.798
ECG 5 000	0.925	0.924	0.941	0.943	<b>0.946</b>	0.945	<b>0.946</b>	0.941	0.931	0.942
ECGFiveDays	0.797	0.768	<b>1.000</b>	0.996	0.999	0.989	<b>1.000</b>	<b>1.000</b>	0.955	0.963
FacesUCR	0.769	0.905	0.957	0.891	0.942	0.929	0.945	<b>0.976</b>	0.958	0.946
Fish	0.783	0.823	0.989	0.983	0.983	0.983	0.966	0.971	0.989	<b>0.994</b>
FordA	0.665	0.555	0.930	0.953	<b>0.957</b>	0.928	0.929	0.951	0.928	0.947
GunPoint	0.913	0.907	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.993	0.993	0.993	<b>1.000</b>
Ham	0.600	0.467	0.667	0.705	0.648	0.791	0.800	<b>0.829</b>	0.781	0.791
HandOutlines	0.862	0.881	0.903	0.886	0.919	0.887	0.886	0.897	0.861	<b>0.924</b>
InlineSkate	0.342	0.384	0.516	<b>0.567</b>	0.495	0.482	0.442	0.336	0.365	0.491
ItalyPowerDemand	0.955	0.950	0.909	0.945	0.961	0.962	0.967	<b>0.975</b>	0.960	0.967
LargeKitchenApp	0.493	0.795	0.765	0.675	0.845	0.910	0.907	0.917	0.893	<b>0.931</b>
Lightning 2	0.754	<b>0.869</b>	0.836	0.721	<b>0.869</b>	0.803	0.852	0.852	0.754	0.803
Meat	0.933	0.933	0.900	0.850	0.917	0.884	0.950	0.950	<b>1.000</b>	0.950
MedicalImages	0.684	0.737	0.718	0.732	0.758	<b>0.801</b>	0.762	0.791	0.772	0.783
MidPhalOutlineCor	0.766	0.698	0.780	0.763	0.804	<b>0.840</b>	0.788	0.818	0.793	0.832
MoteStrain	0.879	0.835	0.879	<b>0.950</b>	0.937	0.922	0.890	0.899	0.895	0.942
OSULeaf	0.521	0.591	0.955	0.975	0.967	<b>0.996</b>	0.897	0.983	0.979	0.988
PhalangesOutCorrect	0.761	0.728	0.772	0.797	0.770	0.823	0.817	0.815	0.825	<b>0.846</b>
Phoneme	0.109	0.228	0.265	0.313	0.349	0.350	0.245	<b>0.359</b>	0.324	0.346
Plane	0.962	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
ProxPhalOutAgeGrp	0.785	0.805	0.834	0.839	0.854	<b>0.883</b>	0.873	0.844	0.849	0.873
RefrigerationDevices	0.395	0.464	0.499	0.544	0.547	<b>0.579</b>	0.555	0.552	0.528	0.552
ScreenType	0.360	0.397	0.464	0.400	0.547	0.659	0.547	0.613	<b>0.707</b>	0.619
SonyAIBORobSurf 1	0.696	0.725	0.632	0.892	0.845	0.982	0.923	0.895	0.985	<b>0.995</b>

表3 (续)

数据集	ED	DTW	BOSS	WEASEL	COTE	LSTM-FCN	EMAN	AFFNet	ResNet	DRPRFs
SonyAIBORobSurf2	0.859	0.831	0.859	0.940	0.952	<b>0.978</b>	0.919	0.967	0.962	0.972
StarLightCurves	0.849	0.907	0.978	0.976	<b>0.980</b>	0.976	0.968	0.978	0.975	0.979
Strawberry	0.946	0.941	0.976	0.978	0.951	<b>0.987</b>	0.974	0.976	0.958	0.981
SwedishLeaf	0.789	0.792	0.922	0.965	0.955	<b>0.979</b>	0.933	0.976	0.958	<b>0.979</b>
SyntheticControl	0.880	0.993	0.967	0.993	<b>1.000</b>	0.997	0.997	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
ToeSegmentation1	0.680	0.772	0.939	0.974	0.974	0.987	0.961	0.974	0.965	<b>0.996</b>
Trace	0.760	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
TwoLeadECG	0.747	0.905	0.981	0.998	0.993	0.999	0.993	0.998	<b>1.000</b>	0.993
TwoPatterns	0.907	<b>1.000</b>	0.993	0.989	<b>1.000</b>	0.997	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
UwaveGestLibZ	0.650	0.658	0.695	0.738	0.750	<b>0.797</b>	0.760	0.791	0.755	0.728
Wafer	0.995	0.980	0.995	0.999	<b>1.000</b>	0.999	0.999	0.999	0.997	0.997
Worms	0.455	0.584	0.558	0.805	0.623	0.702	0.630	<b>0.831</b>	0.619	0.805
WormsTwoClass	0.610	0.623	0.831	0.818	0.805	0.785	0.757	<b>0.857</b>	0.735	0.844
Yoga	0.830	0.837	<b>0.918</b>	0.885	0.877	0.911	0.874	0.879	0.858	0.898
平均值	0.727	0.772	0.837	0.853	0.859	0.881	0.862	0.878	0.866	<b>0.885</b>
胜出率	0.022	0.111	0.133	0.155	0.244	0.311	0.156	0.289	0.178	<b>0.333</b>

表4 MTS分类实验结果对比

数据集	ED	DTW <sub>D</sub>	WEASEL+MUSE	MLSTM-FCN	TapNet	ShapeNet	ResNet	DRPRFs
ArticularyWordRecognition	0.970	0.987	0.990	0.973	0.987	0.987	0.977	<b>0.993</b>
BasicMotions	0.675	0.975	<b>1.000</b>	0.950	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
CharacterTrajectories	0.964	0.989	0.990	0.985	<b>0.997</b>	0.980	0.994	0.992
Cricket	0.944	<b>1.000</b>	<b>1.000</b>	0.917	0.958	0.986	<b>1.000</b>	<b>1.000</b>
Epilepsy	0.666	0.964	<b>1.000</b>	0.761	0.971	0.987	0.978	0.994
ERing	0.133	0.133	0.133	0.133	0.133	0.133	0.133	<b>0.878</b>
FingerMovements	0.550	0.530	0.490	0.580	0.530	0.580	0.560	<b>0.590</b>
HandMovementDirection	0.278	0.231	0.365	0.365	0.378	0.338	<b>0.405</b>	0.311
Handwriting	0.200	0.286	0.605	0.286	0.357	0.451	<b>0.629</b>	0.566
JapaneseVowels	0.924	0.949	0.973	0.976	0.965	0.984	0.957	<b>0.995</b>
Libras	0.833	0.870	0.878	0.856	0.850	0.856	0.956	<b>0.983</b>
LSST	0.456	0.551	0.590	0.373	0.568	0.590	<b>0.634</b>	0.561
NATOPS	0.850	0.883	0.870	0.889	0.939	0.883	0.972	<b>0.994</b>
PenDigits	0.973	0.977	0.948	0.978	0.980	0.977	0.985	<b>0.992</b>
Phoneme	0.104	0.151	0.190	0.110	0.175	0.298	<b>0.349</b>	0.298
RacketSports	0.868	0.803	<b>0.934</b>	0.803	0.868	0.882	0.921	0.927
SelfRegulationSCP 1	0.771	0.775	0.710	<b>0.874</b>	0.652	0.782	0.785	0.724
SelfRegulationSCP 2	0.483	0.539	0.460	0.472	0.550	<b>0.572</b>	0.428	0.461
StandWalkJump	0.200	0.200	0.333	0.067	0.400	<b>0.533</b>	0.267	0.400
UWaveGestureLibrary	0.881	0.903	<b>0.916</b>	0.891	0.894	0.906	0.866	0.863
平均值	0.636	0.685	0.719	0.662	0.708	0.735	0.740	<b>0.776</b>
胜出率	0.000	0.050	0.250	0.050	0.100	0.150	0.300	<b>0.450</b>

此外,本文采用Nemenyi检验来衡量所提出模型与其他分类模型在精度上是否存在显著性差异,在显著水平 $\alpha = 0.05$ 的条件下绘制出临界差异图,如图5所示. 差异图分布了各算法的平均等级,被同一水平线覆盖下的分类器没有统计学上的显著性差异,由

此可以看出,DRPRFs在不需要特征工程的繁琐处理下,比传统方法存在显著的效果提升,尽管与COTE、LSTM-FCN、EMAN、AFFNet、ResNet间的差异不显著,但是,DRPRFs仍然取得了排名第1的平均秩3.933.

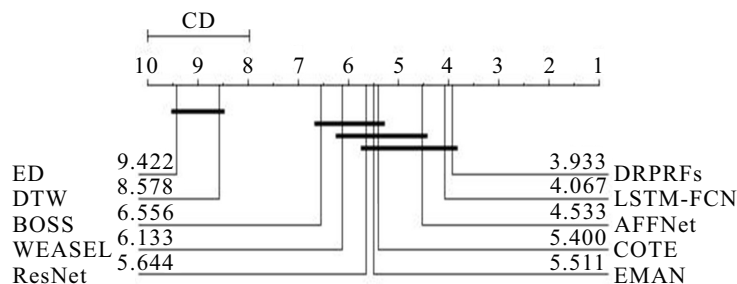


图5 对比模型的临界差异

为了直观地比较所提出方法在 ResNet 的基础上分类性能提升的显著性,图 6 给出了 ResNet 与 DRPRFs 在 65 个 TSC 数据集上的精度对比散点图. 每个点为两个分类器在一个数据集上的分类精度,对角线下方的点表示 DRPRFs 在对应数据集的精度比竞争方法高. 大部分精度散点均偏向 DRPRFs 模型,在 42 个数据集上的性能优于 ResNet,表明所提出模型可提升 ResNet 在 TSC 任务上的分类效果,这主要得益于随机森林是比全连接层更可靠的分类器,且“随机子空间”是可获得更佳泛化能力的有效策略.

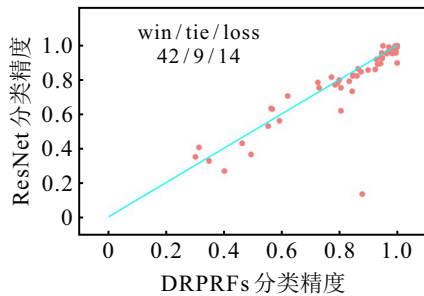
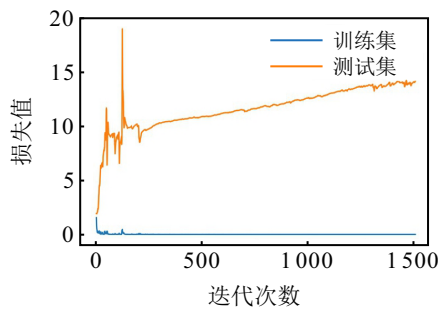


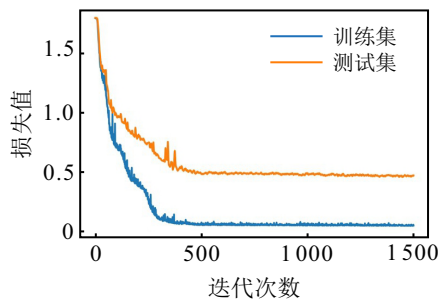
图 6 DRPRFs 与 ResNet 精度对比散点图

### 3.1.4 泛化分析

ResNet 的局限在于其深层网络结构导致在对小规模的时间序列进行分类时易过拟合,而 DRPRFs 在 ResNet 的基础上融入概率随机森林来提高泛化性能. 图 7 为 ResNet 和 DRPRFs 在 ERing 上的训练和测试损失曲线,该数据集仅有 30 个训练样本. 随着训练损失的下降,ResNet 的测试损失值不断上升至 10 以上,产生过拟合现象;而 DRPRFs 的测试损失不断下降并收敛于 0.5 左右,很好地提高了网络的泛化能力.



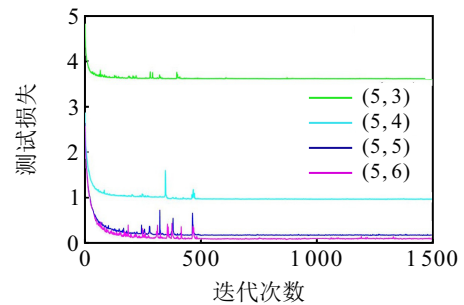
(a) ResNet



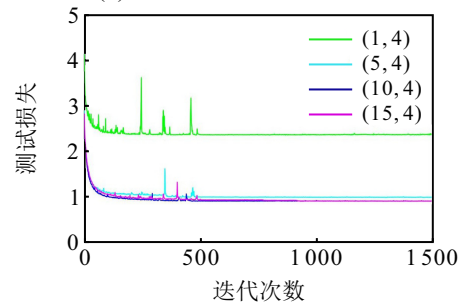
(b) DRPRFs

图 7 数据集 ERing 的训练和测试损失

基分类器的数量和基分类器性能的强弱对于集成模型的性能尤为重要. 通过 SwedishLeaf 来分析树的数量和深度对模型性能的影响,如图 8(a) 所示:控制树的数量为 5, DRPRFs 的测试损失随着树的深度不断增加而降低,这是由于每棵概率决策树学习到的特征数  $(2^h - 1)$  随着树的深度  $h$  呈指数增加,从而提升了基分类器的性能. 图 8(b) 给出了不同树的数量下 DRPRFs 的测试损失,在树的深度为 4 的前提下,增加树的数量会改善 DRPRFs 的分类性能. 过少的树导致集成模型所包含的特征信息不足,减弱了分类性能. 因此,所提出集成策略可通过改变树的数量和深度有效地提升模型在测试集上的泛化表现.



(a) 树的深度对模型性能的影响



(b) 树的数量对模型性能的影响

图 8 模型泛化表现的比较分析

### 3.1.5 模型部件验证

为了验证 ResNet 和概率随机森林 (probability random forests, PRFs) 在模型中的有效性和作用,本节将文献 [9] 中的 MLP 和 FCN 与概率随机森林相结合,记为 MP 和 FP,分别在 BirdChicken 等 6 个数据集上进行实验,实验结果如表 5 所示.

表 5 模型部件验证实验结果

数据集	MLP	FCN	ResNet	MP	FP	DRPRFs
B.C	0.800	0.950	0.900	0.600	0.950	<b>1.000</b>
Fish	0.874	0.971	0.989	0.880	0.931	<b>0.994</b>
I.S	0.351	0.411	0.365	0.313	0.422	<b>0.491</b>
M.I	0.729	<b>0.792</b>	0.772	0.539	0.758	0.783
S.A.R.S	0.727	0.968	0.985	0.651	0.945	<b>0.995</b>
W.T.C	0.597	0.729	0.735	0.636	0.779	<b>0.844</b>

分别对比 MP 与 MLP、FP 与 FCN 以及 DRPRFs 与 ResNet,发现 3 种网络在融入 PRFs 后分类精度大多有所提升,体现了 PRFs 的良好分类性能. 相较于 MP 与

FP, DRPRFs在5个数据集上取得了最佳效果,这主要归功于ResNet比MLP、FCN这类浅层网络模型更能够挖掘出原始时间序列所蕴含的深层特征信息. 在BirdChick和InlineSkate这两个训练集规模很小的数据集上, ResNet由于过拟合问题表现出劣于FCN的分类性能,但是,PRFs的融入成功地提升了深度网络的泛化能力,使得DRPRFs在这两个数据集上取得了优于FP的分类精度.

### 3.2 液压系统的故障诊断

本节将所提出模型应用于液压系统的故障诊断这一工业案例,以进一步验证DRPRFs模型能够有效解决实际工业生产中多变量时间序列分类问题.

液压状态监测系统 (hydraulic condition monitoring system, HCMS) 是一类非线性、多变量的复杂系统,具有多工况时空耦合、多参数监测信号等特征,因此,难以获取其故障机理. HCMS包含能源和伺服两个子系统,冷却器、液压泵、阀门和蓄能器是保证系统稳定可靠运行的核心部件,诊断关键元件的故障类型对于维护液压系统具有重要的现实意义. 随着传感器技术的发展,HCMS保存了大量的液压系统状态监测数据,因此,可基于HCMS的多元时序数据挖掘故障特征与故障类型的映射关系. 整个系统的传感器网络由压力、流量、温度、冷却功率和电机功率等17个传感器组成,采样频率范围为100 Hz~1 Hz. 尽管传感器收集到的时序信号在一定程度上根据元件的故障类型表现出不同的表现形式,部分时序信号在不同故障类型下则没有明显的变化,图9展示出压力传感器PS1对蓄能器不同严重程度的故障无明显差异特征,根据单一时序信号的表现特征是难以确定故障类型的,因此,需要综合考虑整个传感器网络的数据信息.

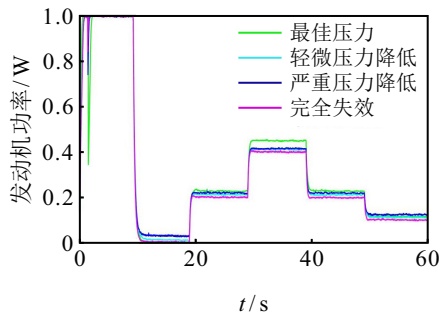


图9 不同故障类型的PS1时间序列表现

本文以UCI机器学习库中的液压系统状态监测数据<sup>[25]</sup>为数据来源,分别对冷却器、液压泵、阀门和蓄能器这4个核心元件进行故障诊断实验,其故障信息如表6所示. 考虑到传感器间采样速率和单位量纲的不一致,本文对17个传感器的原始时序数据进

行重采样和归一化预处理. 预处理的步骤包括:首先向上采样到最大采样率100 Hz,然后通过传感器的每6个连续读数取平均值来进行降采样,同时将传感器数据归一化到0~1范围内. 将2205条故障数据按7:3的比例随机划分为训练集和测试集,其中每条故障数据包含17维时间序列长度为1000的时序数据. 在故障诊断实验中,对比方法包括文献[26]的卷积神经网络(CNN)、文献[27]的门控卷积自编码器(GCAEC)、MLSTM-FCN和ResNet. 实验结果如表7所示. DRPRFs在液压泵、阀门和蓄能器的故障模式识别中取得了最优精度. DRPRFs在HCMS的关键元件的诊断率接近100%,很好地满足了实际工业生产中的故障诊断问题.

表6 液压系统故障信息

故障元件	故障状态	样本数
冷却器	完全失效	732
	效率降低	732
	正常工作	741
液压泵	严重泄漏	492
	轻微泄漏	492
	无泄漏	1221
阀门	完全失控	360
	严重延迟	360
	轻微延迟	360
	状态切换正常	1125
蓄能器	几乎完全失效	808
	严重压力降低	399
	轻微压力降低	399
	最佳压力	599

表7 故障诊断精度对比

故障元件	CNN	GCAEC	MLSTM-FCN	ResNet	DRPRFs
冷却器	0.996	<b>1.000</b>	0.998	0.998	0.998
液压泵	0.969	0.995	0.991	<b>0.998</b>	<b>0.998</b>
阀门	<b>1.000</b>	<b>1.000</b>	0.994	<b>1.000</b>	<b>1.000</b>
蓄能器	0.982	0.977	0.664	0.939	<b>0.998</b>

## 4 结论

针对当前时间序列分类研究存在的问题,本文提出了一种新的端到端的时间序列分类深度模型,有效地结合了残差网络的特征提取能力以及随机森林的拟合数据和泛化能力. 该模型利用残差网络的深层网络结构提取原始时间序列数据的高级判别性特征,并将概率随机森林融入至网络的分类层,更好地建立起数据特征与标签间的映射关系. 通过定义一个全局目标函数,残差网络和随机森林的结构参数可以以反向传播的方式得到联合优化. 利用“随机子空间”的集成策略缓解了深度神经网络在时间序列分类任务中的过拟合现象,提高了深度结构的泛化能力. 大量实验表明,DRPRFs具有显著的分类性能,同时在小规模时序数据集上具有较好的泛化能力.

## 参考文献(References)

- [1] Li H L, Liang Y, Wang S C. Review on dynamic time warping in time series data mining[J]. *Control and Decision*, 2018, 33(8): 1345-1353.
- [2] Li H, Wu X, Wan X, et al. Time series clustering via matrix profile and community detection[J]. *Advanced Engineering Informatics*, 2022, 54: 101771.
- [3] Yuan J D, Wang Z H, Han M. Shapelet pruning and shapelet coverage for time series classification[J]. *Journal of Software*, 2015, 26(9): 2311-2325.
- [4] Schäfer P. The BOSS is concerned with time series classification in the presence of noise[J]. *Data Mining and Knowledge Discovery*, 2015, 29(6): 1505-1530.
- [5] Li H L, Liang Y. Similarity measure based on numerical symbolic and shape feature for time series[J]. *Control and Decision*, 2017, 32(3): 451-458.
- [6] Li H L, Jia R Y, Wan X J. Time series classification based on complex network[J]. *Expert Systems with Applications*, 2022, 194: 116502.
- [7] Bagnall A, Lines J, Hills J, et al. Time-series classification with COTE: The collective of transformation-based ensembles[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(9): 2522-2535.
- [8] Lines J, Taylor S, Bagnall A. Time series classification with HIVE-COTE: The hierarchical vote collective of transformation-based ensembles[J]. *ACM Transactions Knowledge Discovery from Data*, 2018, 12(5): 1-36.
- [9] Wang Z G, Yan W Z, Oates T. Time series classification from scratch with deep neural networks: A strong baseline[C]. *International Joint Conference on Neural Networks*. Anchorage, 2017: 1578-1585.
- [10] Karim F, Majumdar S, Darabi H, et al. LSTM fully convolutional networks for time series classification[J]. *IEEE Access*, 2017, 6: 1662-1669.
- [11] Ismail F H, Forestier G, Weber J, et al. Deep learning for time series classification: A review[J]. *Data Mining and Knowledge Discovery*, 2019, 33(4): 917-963.
- [12] Ma Q L, Zheng Z J, Zhuang W Q, et al. Echo memory-augmented network for time series classification[J]. *Neural Networks*, 2021, 133: 177-192.
- [13] Wang T, Liu Z Y, Zhang T, et al. Adaptive feature fusion for time series classification[J]. *Knowledge-Based Systems*, 2022, 243: 108459.
- [14] Zhang Y W, Wang Z H, Liu H Y, et al. Time series classification algorithm based on multiscale residual full convolutional neural network[J]. *Journal of Software*, 2022, 33(2): 555-570.
- [15] Kotschieder P, Fiterau M, Criminisi A, et al. Deep neural decision forests[C]. *IEEE International Conference on Computer Vision*. Santiago, 2015: 1467-1475.
- [16] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, 2016: 770-778.
- [17] Dau H A, Bagnall A, Kamgar K, et al. The UCR time series archive[J]. *IEEE/CAA Journal of Automatica Sinica*, 2019, 6(6): 1293-1305.
- [18] Bagnall A, Dau H A, Lines J, et al. The UEA multivariate time series classification archive[J/OL]. 2018, arXiv: 1811.00075.
- [19] Schäfer P, Leser U. Fast and accurate time series classification with weasel[C]. *Proceedings of the ACM on Conference on Information and Knowledge Management*. Singapore, 2017: 637-646.
- [20] Shokoohi-Yekta M, Wang J, Keogh E. On the non-trivial generalization of dynamic time warping to the multi-dimensional case[C]. *Proceedings of the SIAM International Conference on Data Mining*. Philadelphia, 2015: 289-297.
- [21] Schäfer P, Leser U. Multivariate time series classification with WEASEL+MUSE[J/OL]. 2017, arXiv: 1711.11343.
- [22] Karim F, Majumdar S, Darabi H, et al. Multivariate LSTM-FCNs for time series classification[J]. *Neural Networks*, 2019, 116: 237-245.
- [23] Zhang X C, Gao Y F, Lin J, et al. TapNet: Multivariate time series classification with attentional prototypical network[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(4): 6845-6852.
- [24] Li G Z, Choi B, Xu J L, et al. ShapeNet: A shapelet-neural network approach for multivariate time series classification[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(9): 8375-8383.
- [25] Helwig N, Pignanelli E, Schütze A. Condition monitoring of a complex hydraulic system using multivariate statistics[C]. *IEEE International Instrumentation and Measurement Technology Conference Proceedings*. Pisa, 2015: 210-215.
- [26] König C, Helmi A M. Sensitivity analysis of sensors in a hydraulic condition monitoring system using CNN models[J]. *Sensors*, 2020, 20(11): 3307.
- [27] Gareev A, Protsenko V, Stadnik D, et al. Improved fault diagnosis in hydraulic systems with gated convolutional autoencoder and partially simulated data[J]. *Sensors*, 2021, 21(13): 4410.

## 作者简介

刘颖(1982-), 女, 副教授, 博士, 从事一体化生产计划与调度、系统仿真建模智能优化算法及应用等研究, E-mail: liu\_ying@dlut.edu.cn;

李旭(1997-), 男, 硕士生, 从事时间序列分类、深度学习等研究, E-mail: li\_xu@mail.dlut.edu.cn;

吕政(1987-), 男, 副教授, 博士, 从事机器学习、复杂系统调度、知识自动化等研究, E-mail: lvzheng@dlut.edu.cn;

赵珺(1981-), 男, 教授, 博士生导师, 从事流程工业生产计划与优化调度、计算机集成制造等研究, E-mail: zhaoj@dlut.edu.cn;

王伟(1955-), 男, 教授, 博士生导师, 从事复杂系统建模控制与优化、流程工业生产计划与优化调度等研究, E-mail: wangwei@dlut.edu.cn.