

## 基于异构特征融合的多维时间序列分类算法

乔帆, 王鹏, 汪卫

引用本文

乔帆, 王鹏, 汪卫. [基于异构特征融合的多维时间序列分类算法](#)[J]. 计算机科学, 2024, 51(2): 36-46.

QIAO Fan, WANG Peng, WANG Wei. [Multivariate Time Series Classification Algorithm Based on Heterogeneous Feature Fusion](#) [J]. Computer Science, 2024, 51(2): 36-46.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

### [基于改进自注意力机制和表示学习的分层文档分类方法](#)

Hierarchical Document Classification Method Based on Improved Self-attention Mechanism and Representation Learning

计算机科学, 2024, 51(2): 238-244. <https://doi.org/10.11896/jsjcx.221100266>

### [基于全局与局部特征的二阶段文档图像可见水印去除模型](#)

Two-stage Visible Watermark Removal Model Based on Global and Local Features for Document Images

计算机科学, 2024, 51(2): 172-181. <https://doi.org/10.11896/jsjcx.230600144>

### [基于扩张卷积条件生成对抗网络的红外小目标检测](#)

Infrared Small Target Detection Based on Dilated Convolutional Conditional Generative Adversarial Networks

计算机科学, 2024, 51(2): 151-160. <https://doi.org/10.11896/jsjcx.221200045>

### [基于特征融合与边界修正显著性目标检测](#)

Feature Fusion and Boundary Correction Network for Salient Object Detection

计算机科学, 2023, 50(12): 166-174. <https://doi.org/10.11896/jsjcx.221100203>

### [基于Transformer特征融合的时间序列分类网络](#)

Transformer Feature Fusion Network for Time Series Classification

计算机科学, 2023, 50(12): 97-103. <https://doi.org/10.11896/jsjcx.221100112>

# 基于异构特征融合的多维时间序列分类算法

乔帆<sup>1</sup> 王鹏<sup>2</sup> 汪卫<sup>2</sup>

<sup>1</sup> 复旦大学软件学院 上海 200438

<sup>2</sup> 复旦大学计算机科学与技术学院 上海 200438

(fqiao20@fudan.edu.cn)

**摘要** 随着大数据时代的到来和传感器的发展,多维时间序列分类问题成为数据挖掘领域的重要问题。多维时间序列存在维度高、维度间关系复杂、数据形态多变的特点,从而生成巨大的特征空间。现有方法难以选取有区分力的特征,导致方法的准确度普遍较低。另一方面,现有方法的分类结果的可解释性较差。针对上述问题,提出了一种基于异构特征融合的多维时间序列分类算法。该算法融合了时域、频域和区间统计值这3种特征并对特征进行聚类,从而找到最有代表性的特征。首先为每个维度提取不同类型的代表性特征,再通过多维度特征转换的方法融合所有维度的不同类型的特征,形成特征向量,并基于此训练分类模型。为了提高分类结果的可解释性,算法基于树结构生成不同类型的候选特征集合,然后通过聚合消除冗余和相似的特征,最终获得少量代表性特征。为了验证所提算法的有效性,在公开的UEA数据集上进行了大量实验。实验结果显示,所提算法的准确性、特征融合的合理性,以及分类结果的可解释性均优于现有方法。

**关键词**: 多维度时间序列; 时间序列分类; 特征融合; 可解释性; 特征聚类

中图分类号 TP391

## Multivariate Time Series Classification Algorithm Based on Heterogeneous Feature Fusion

QIAO Fan<sup>1</sup>, WANG Peng<sup>2</sup> and WANG Wei<sup>2</sup>

<sup>1</sup> School of Software, Fudan University, Shanghai 200438, China

<sup>2</sup> School of Computer Science, Fudan University, Shanghai 200438, China

**Abstract** With the advance of big data and sensors, multivariable time series classification has been an important problem in data mining. Multivariate time series are characterized by high dimensionality, complex inter-dimensional relations, and variable data forms, which makes the classification methods generate huge feature spaces, and it is difficult to select discriminative features, resulting in low accuracy and hindering the interpretability. Therefore, a multivariate time series classification algorithm based on heterogeneous feature fusion is proposed in this paper. The proposed algorithm integrates time-domain, frequency-domain, and interval-based features. Firstly, a small number of representative features of different types are extracted for each dimension. Then, features of all dimensions are fused by multivariable feature transformation to learn the classifier. For univariate feature extraction, the algorithm generates different types of feature candidates based on tree structure, and then a clustering algorithm is designed to aggregate redundant and similar features to obtain a small number of representative features, which effectively reduces the number of features and enhances the interpretation of the method. In order to verify the effectiveness of the algorithm, expensive experiments are conducted on the public UEA dataset, and the proposed algorithm is compared with the existing multivariate time series classification methods. The results prove that the proposed algorithm is more accurate than the comparison methods, and the feature fusion is reasonable. What's more, the interpretability of classification results is showed by case study.

**Keywords** Multivariate time series, Time series classification, Feature fusion, Interpretability, Feature clustering

## 1 引言

时间序列是随时间变化采集得到的一组等间隔的有序实值序列,在金融<sup>[1]</sup>、医学<sup>[2]</sup>、遥感<sup>[3]</sup>等领域被广泛应用。根据时间戳的变量是否单一,可将时间序列分为单维时间序列

(Univariate Time Series, UTS)和多维时间序列(Multivariate Time Series, MTS)。多维时间序列指同一时间内对多个维度同时采集得到的一组同步有序观测数值,脑电图(Electroencephalography, EEG)和脑磁图(Magnetoencephalography, MEG)是典型的多维时序数据。

到稿日期:2023-01-31 返修日期:2023-05-16

基金项目:科技部重点研发计划(2020YFB1710001)

The work was supported by the Key Research and Development Program of Ministry of Science and Technology of China(2020YFB1710001).

通信作者:王鹏(pengwang5@fudan.edu.cn)

时间序列分类(Time Series Classification, TSC)作为数据挖掘的重要任务,随着大数据时代和工业 4.0 的发展,越来越受到人们的重视。时序分类方法主要分为基于形状、基于结构、基于区间、基于集成和深度学习的方法。基于形状的方法通过度量原始时间序列的相似度,构建分类模型。最近邻模型(1-NN)采用欧氏距离或动态时间规整(Dynamic Time Warping, DTW)<sup>[4]</sup>距离度量序列之间的相似性。2009年,经典的基于 shapelet 的分类方法<sup>[5]</sup>首次被提出。shapelet 是时间序列中有区分能力的一段连续子序列,本身具有可解释性, SAX-VFSEQL<sup>[6]</sup>, ShapeNet<sup>[7]</sup>等方法则对其进行了扩展。但是由于大多数方法需要生成大量 shapelet,因此分类算法的可解释性大幅降低。基于结构的方法对原始时间序列进行转换,得到特征向量。常见特征有两类:一类是统计特征,如均值、标准差、自相关系数;另一类是符号表示,例如符号聚合近似(Symbolic Aggregate Approximation, SAX)<sup>[8]</sup>和符号傅里叶近似(Symbolic Fourier Approximation, SFA)<sup>[9]</sup>。基于区间的方法提取时间序列位于固定区间的特征,当数据存在大量噪声时,该类方法可以排除噪声引起的波动干扰,定位有效区间。TSF<sup>[10]</sup>和 RISE<sup>[11]</sup>是该类方法的典型代表。基于集成的方法组合不同类型的特征或分类器,以达到提高准确度的目的, HIVE-COTE2.0<sup>[12]</sup>, TS-CHIEF<sup>[13]</sup>都是优秀的集成方法。但上述方法普遍存在特征数量过多致使方法可解释性降低的问题,尤其是集成方法,虽然提高了准确度,但是多个分类模型生成不同类型的大量特征使得分类结果可解释性很弱。近年来,深度学习在时间序列分类上得以应用<sup>[14]</sup>,网络的结构使其易于扩展更多的维度,因此诸多多维时序分类的研究聚焦于深度学习。ResNet<sup>[15]</sup>由 3 个卷积层组成,2017 年首次被用于时间序列领域,可用于单变量和多变量分类;时间序列注意力原型网络(Time series attentional prototype Network, TapNet)<sup>[16]</sup>是专用于多维时序分类的神经网络,它通过网络学习时间序列的潜在特征;基于 shapelet 的神经网络 ShapeNet<sup>[7]</sup>通过神经网络的方式将不同长度的子序列嵌入到统一空间,最后使用标准分类器分类。广义随机 shapelet 森林(Generalized Random Shapelet Forests, gRSF)<sup>[17]</sup>生成一组基于 shapelet 的决策树,在每个节点随机选择一个维度进行处理。然而,深度学习的“黑盒”特性使得分类结果更难具有可解释性。

现有的时间序列分类主要聚焦于单维时间序列分类(Univariate Time Series Classification, UTSC),对多维时序分类(Multivariate Time Series Classification, MTSC)的研究较少。多维时间序列维度高、维度间数据关联依赖的关系复杂,同时还存在大量不相关的数据和噪音。多维度数据的特点使得传统方法生成非常大的特征空间,例如 shapelet 或符号单词,极大地提高了特征选择的复杂性,有区分力的特征被掩盖,无法获取期望的准确度,结果的可解释性也无从考量。深度学习方法虽然适用于多维数据,然而其对样本个数和计算资源的要求较高,且结果的可解释性很低,阻碍了方法的进一步调优与结果的有效分析应用。

综上所述,现有 TSC 方法存在准确度与可解释性难以兼得的问题。MTSC 面临的挑战更为严峻:1)大规模的特征

空间增加了有区分力特征选择的难度,导致结果不够准确;2)分类方法的种类与参数难以调整,因此无法适用于不同复杂场景下多维度的时间序列数据;3)大量特征和深度学习方法的“黑盒”训练过程都无法提供可解释的分类结果。

针对上述问题,本文提出了基于异构特征融合的多维时间序列分类算法,简称 HFC。本文的主要贡献如下:

1)提出了异构特征的通用定义。集成方法可解释性低的一个主要原因是,不同方法对特征的概念有不同的理解,无法对特征统一化。本文抽象出用于分类的不同类型特征的通用定义,便于对特征进行处理,同时使得特征具有可解释性。

2)提出了基于树结构的异构特征生成方法。算法基于类似决策树的机制,为每个维度生成高质量的时域特征、频域特征和区间特征,构建异构特征空间,增强方法的普适性,通过方法参数的随机选择,增强方法的鲁棒性。

3)提出了基于聚类的代表性特征的提取方法。针对不同类型的特征,设计聚类方法以去除大量冗余和相似的特征,为每个维度提取少量代表性特征,极大地压缩特征空间,提高方法的准确度,使得每个维度的特征可解释。

4)提出了面向多维时序的扩展方法——多维度特征转换(Multivariate Feature Transformation, MFT)方法。对于一个多维时间序列,本文计算每个特征与其来源维度对应数据的距离,将原始多维序列转换为一个距离向量,即一个 MFT 表示。经过 MFT,本文最终得到一个距离矩阵,有效融合多维度的异构特征,便于标准分类器学习分类模型。本文选择线性支持向量机(Support Vector Machine, SVM)作为分类器,使得最终分类结果直观可解释。

## 2 相关工作

MTSC 主要分为两类,一类是通过使用 PCA 等方法降维或串联 MTS 的所有维度数据,将 MTS 转换为 UTS,以及通过组合 UTSC 模型的方式,将 MTSC 问题转换为 UTSC 问题,该类方法使用现有的大量 UTSC 方法即可;另一类是专为 MTSC 定制的方法,包括将 UTSC 扩展到 MTS,本节主要讨论这种情况。

1NN-DTW<sup>[18]</sup>作为最基本的 MTSC 方法,包括 DTW<sub>1</sub>和 DTW<sub>D</sub>两个版本。DTW<sub>1</sub>简单地将每个维度的单维 DTW 距离相加,DTW<sub>D</sub>计算动态规划两步之间的距离时,考虑所有维度的两个向量间的欧氏距离。gRSF<sup>[17]</sup>生成集成基于 shapelet 的决策树,对于多维的情况,在每个节点处随机选择一个维度作为生成 shapelet 的数据。超速 shapelet(Ultra Fast Shapelets, UFS)<sup>[19]</sup>使用随机的超快速 shapelet,从多维时间序列中有效选择代表性的模式进行分类。模式相似性学习(Learned Pattern Similarity, LPS)<sup>[20]</sup>从一个多维时间序列中提取片段,训练回归树识别片段之间的依赖性,然后为叶子节点的标签建立词袋模型。多维度时间序列符号表示(Symbolic Representation for Multivariate Time series, SMTS)<sup>[21]</sup>使用一个编码集捕捉不同维度之间的局部关系,对多维度时间序列进行分类。WEASEL-MUSE<sup>[22]</sup>在序列原始维度和每个维度相邻两点差值构成的新维度上提取长度不同的 SFA 单词,通过直方图统计个数,使用卡方检验进行特征选择。典型

区间森林(Canonical Interval Forest, CIF)<sup>[23]</sup>随机选择  $k$  个区间,使用 Catch22<sup>[24]</sup>特征和统计特征,构建时序树的集合。多重表征的典型区间森林分类器(Diverse Representation Canonical Interval Forest Classifier, DrCIF)<sup>[12]</sup>扩展 CIF,添加周期图和差分表示特征。对于 MTSC 问题,随机选择维度并计算特征。

深度学习由于网络对维度具有可扩展性,因此能够解决 MTSC 问题。多维长时记忆全卷积网络(Multivariate Long Short Term Memory-Fully Convolutional Network, MLSTM-FCN)<sup>[25]</sup>使用 LSTM 和连续的 CNN 层提取特征后,使用 softmax 层预测分类标签。TapNet<sup>[6]</sup>提出一个带有注意力原型网络的分类模型,学习 MTS 潜在特征。ShapeNet<sup>[7]</sup>将长度不同的 shapelet 嵌入到统一空间,提出基于聚类的损失函数,采用无监督的方式训练网络。多通道深度卷积神经网络(Multi Channel Deep Convolutional Neural Network, MCDC-NN)<sup>[26]</sup>通过单独的 CNN 层处理每个单维度时间序列,然后串联所有单维度时间序列的输出,通过一个 softmax 层预测标签。基于自回归森林的多维度时间序列模型(Autoregressive Forests for Multivariate Time Series Modelling, mvARF)<sup>[27]</sup>在一个自回归模型上训练树的集群,捕捉 MTS 个维度特征之间的线性和非线性关系。基于负采样的多维度时间序列分类(Negative Samples, NS)<sup>[28]</sup>采用无监督方法,结合基于因果扩张卷积的编码器和负采样的三元组损失,学习时间序列的通用表示。随机卷积核变换(Random Convolutional Kernel Transform, ROCKET)<sup>[22]</sup>具有高准确度和低计算负担的特点,通过大量随机卷积核转换时间序列,使用生成的特征训练线性分类器,在 26 个 UEA 数据集上取得了最佳评估效果<sup>[15]</sup>。sktime 软件包<sup>1)</sup>对 ROCKET 实现了 MTSC 的扩展。MiniRocket<sup>[23]</sup>对 ROCKET 进一步改进,在保持准确度的同时能够以更快的速度处理更大的数据集。深度学习方法通过训练网络学习端到端模型,无法提供分类结果的可解释性。ShapeNet 虽然通过 shapelet 提供了一定的可解释性,但是其只考虑时域特征,且训练效率低,对计算性能有一定要求。

### 3 前置知识与方法概述

**定义 1(单维度时间序列)**  $T = (t_1, t_2, \dots, t_n)$  表示相等时间间隔的有序实值序列,长度为  $n$ 。

**定义 2(子序列)** 时间序列中一段由连续值组成的长度为  $l$  的数值序列  $(t_i, t_{i+1}, \dots, t_{i+l})$ ,表示为  $T_{i:l}$  或  $T[i, i-l+1]$ 。

**定义 3(多维度时间序列)**  $T = (t_1, t_2, \dots, t_m)$  表示维度为  $m$  的一组有序实值序列,  $t_i = (t_{i,1}, t_{i,2}, \dots, t_{i,n})$ 。

**定义 4(数据集)** 样本个数为  $N$  的 MTS 数据集表示为  $D = \{(T_1, L_1), (T_2, L_2), \dots, (T_N, L_N)\}$ ,其中,  $T_i$  表示维度为  $m$  的多维时间序列,每个维度的数据是长度为  $n$  的单位时间序列;  $L_i$  表示序列  $T_i$  的类别标签,  $L_i \in \{c_1, c_2, \dots, c_{|C|}\}$ ;  $|C|$  表示包含于数据集  $D$  中的类别总数。

信息增益用于评估特征将整个数据集划分成两个原始类别数据集的能力,信息增益越大,表示特征携带的信息越多,

特征越重要。信息增益的计算一般通过信息熵的变化程度来衡量<sup>[5]</sup>,下面分别给出信息熵和信息增益的定义与计算式。

**定义 5(信息熵)** 给定数据集  $D, L_i \in \{c_1, c_2, \dots, c_{|C|}\}$ ,每个类别的数据集合由  $D_{c_i}$  表示。  $p(D_{c_i}) = N_{c_i}/N, N_{c_i}$  表示数据集  $D_{c_i}$  的个数;  $D$  的熵使用  $I(D)$  表示,计算式如下:

$$I(D) = - \sum_{i=1}^{|C|} p(D_{c_i}) \log(p(D_{c_i})) \quad (1)$$

**定义 6(信息增益)** 给定数据集  $D$ ,根据特征  $f$  将  $D$  划分为子集  $D_1$  和  $D_2$ ,两个子集的个数分别为  $N_1$  和  $N_2$ 。划分前后的信息熵分别表示为  $I(D)$  和  $\hat{I}(D)$ ,则划分后的信息熵为:

$$\hat{I}(D) = p(D_1)I(D_1) + p(D_2)I(D_2) \quad (2)$$

根据划分前后的信息熵的差,得到特征  $f$  的信息增益,通过  $gain(f) = I(D) - \hat{I}(D)$  计算。

使用特征划分数据集时,特征与任一数据的距离都可作为划分的阈值,划分后取得信息增益最大的距离阈值称为最佳分隔点。

图 1 给出了本文的方法概述。HFC 采用先分后合的思想解决多维时序分类问题,先提取单个维度的有效特征,再根据不同维度的所有异构特征将一个多维时间序列实例转换为一个特征向量,使得算法既能关注每个维度特性,又能考虑不同维度间的关系。因此, HFC 先将 MTS 数据集按照不同维度进行划分,得到每个维度的序列集合,这些集合相当于单维时序数据集。如图 1 左边部分所示,一个维度为 6 的 MTS 数据集被拆分为 6 个不同的单维时序数据集。特征提取针对这样的单维时间序列展开,由特征生成和特征聚类两步完成。

1)特征生成。其目的是获得不同类型的候选特征集合。本文采用基于类似决策树的树结构,生成由时域、频域和区间 3 种类型的特征组成的候选特征集合  $F_0$ 。多类型特征能够捕捉数据的不同特性,增强算法的普适性,使算法更好地适应于 MTSC 的各种应用场景。对于参数调整的问题,本文采用每个树节点随机选取参数的方式来解决这一问题,增强了方法的鲁棒性。随机性可能带来低质量的特征,例如选择的参数不适用于当前数据集的情况。因此, HFC 使用多个树结构并行生成特征,且每个节点生成多个候选特征。图 1 中的绿色方形表示时域特征,橘色菱形表示频域特征,蓝色圆形表示区间特征,展示了 3 个并行的树结构提取特征的案例。

2)特征聚类。其目的是缩小特征空间,提取少量高质量的代表性特征,增强准确度和可解释性。尽管特征生成时在节点进行特征选择,但多个并行结构以及每个节点多次生成候选特征依然使得候选特征集合  $F_0$  数量巨大,同时包含随机性带来的低质量特征。因此,如图 1 所示,特征聚类步骤中,首先基于特征的信息增益对  $F_0$  进行筛选,去除特征生成阶段的低质量特征,获取最佳特征组合  $F$ 。通过为不同类型的特征设计距离计算方法,并选择最适合的聚类过程,算法进一步聚合  $F$  中的冗余和相似特征,得到单个维度最有代表性的特征  $RF$ 。少量代表性特征不仅能提高分类准确性,还能结果为结果赋予可解释的意义。

<sup>1)</sup> [https://www.sktime.org/en/stable/api\\_reference/auto\\_generated/sktime.classification.kernel\\_based.RocketClassifier.html](https://www.sktime.org/en/stable/api_reference/auto_generated/sktime.classification.kernel_based.RocketClassifier.html)

至此,HFC得到了每个维度的代表性特征。对于多维时间序列,本文提出多维度特征转换方法,融合不同维度提取到的异构特征集合。MFT计算每个多维时间序列与所有特征的距离,将一个多维数据转换为一个距离向量。最终,如图1

中步骤3模块所示,所有的距离向量组合,将多维度时间序列数据集转换为一个距离矩阵,易于学习各种分类模型。本文采用线性SVM作为最终分类器,训练权重表示特征的重要性,便于分类结果特征的可视化,进一步增强可解释性。

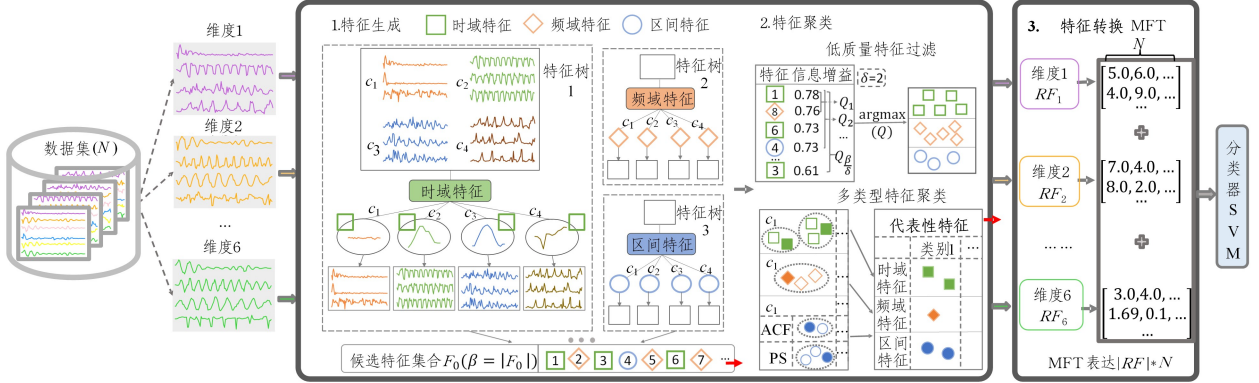


图1 基于异构特征融合的多维时间序列分类算法概述(电子版为彩图)

Fig. 1 Overview of multivariate time series classification algorithm based on heterogeneous feature fusion

### 4 基于异构特征融合的多维时序分类算法

本章详细阐述方法的细节实现,首先介绍3种类型的特征,然后介绍单维度的特征生成与聚类,最后介绍多维度特征转换方法,以及如何配合SVM分类器得到分类结果。

#### 4.1 特征定义

##### 4.1.1 特征的概念与通用定义

特征通常由一段有区分力的子序列的形状或其统计特征表示,图2给出了特征的直观示意图。左侧为时域上的特征挖掘,如图所示,类别A相比B存在一段弧形突起的子序列(红色加粗片段),因此,分类方法可以根据时间序列是否存在与该子序列相似的形状,判断其是否属于类别A。这样的子序列就是时域特征的表现形式,反映了时域上数据的波动形态。右侧部分展示了频域特征的表现形式,可以看到,数据在时域部分并无明显的特征,然而根据傅里叶变换将其转换到频域时,类别B在频域的波形图中存在一个尖形突起。因此,该段子序列作为频域特征的表现形式,可以识别时间序列是否属于类别B。考虑到计算与存储效率,本文采用符号形式存储特征的形状表示。

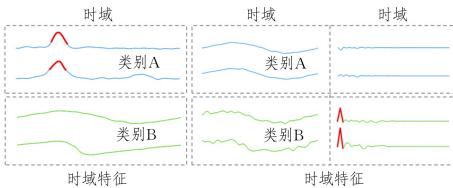


图2 时域特征与频域特征示意图(电子版为彩图)

Fig. 2 Illustration of time domain feature and frequency domain feature

HFC集成时域、频域和区间3种类型的特征,本节首先给出不同类型特征的通用定义,然后依次描述各类特征的具体实现。

**定义7(特征)** 特征以四元组的形式表示为  $f = \langle R_f, P_f, L_f, \theta_f \rangle$ 。特征的表现形式  $R_f$  与生成特征的方法相关,在本文中表现为一个字符型单词或者连续值;  $P_f$  记录方法的

参数;特征标签  $L_f$  表示特征  $f$  代表的类别,即特征所能识别的类;阈值  $\theta_f$  衡量时间序列与特征的相似性,判断数据是否属于类别  $L_f$ 。

上述定义的特征  $f$  识别时间序列  $T$  是否为其代表类别的方式为:首先,在  $T$  上应用生成特征  $f$  的方法,参数和  $P_f$  保持一致,将  $T$  转换为和表现形式  $R_f$  一样的结果;然后,针对不同类型的特征,本文通过计算  $T$  的转换结果和  $R_f$  之间的距离,度量  $T$  与  $f$  的相似性,记为  $D(T, R_f)$ 。  $D(T, R_f) < \theta_f$  时,表示  $f$  识别  $T$  的类别为其代表类别  $L_f$ ,记为  $T \in f$ 。

##### 4.1.2 3种类型特征的具体定义

基于4.1.1节提出的特征通用定义,下文将分别介绍时域特征、频域特征和区间特征的具体定义与生成方式。

时域特征用于捕捉原始序列的形状特征。为了提高计算效率,本文采用SAX转换<sup>[8]</sup>作为时域特征的生成方式。SAX转换的具体方式如下:首先,将标准化的时间序列划分为  $w$  个长度相等的子段,计算每段数据的均值,将序列转换成长度为  $w$  的分段聚合近似(Piecewise Aggregate Approximation, PAA)向量。然后,对PAA值离散化,根据高斯分布分桶,将每个PAA值划分到  $\Sigma$  个区间中,最终将时间序列转变为一个符号单词,称为SAX单词。

相比整个序列,本文在特征提取时更关注于有区分力的子序列。因此,对于时间序列  $T$ ,HFC通过长度为  $l$  的滑动窗口提取子序列,在其上应用SAX转换,最终将整个序列转换为  $|T| - l + 1$  个SAX单词,每个单词对应一个子序列。图3给出了图2中时域特征来源序列的SAX转换示例,选择  $l = 100$  的滑动窗口,圆圈中的序列被转换为 ddedc 的SAX单词。

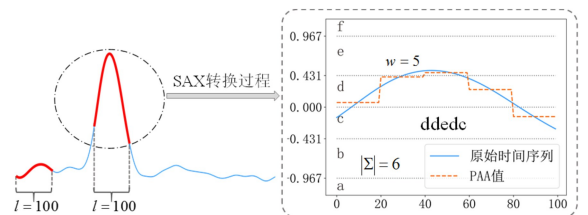


图3 SAX转换过程实例

Fig. 3 Example of SAX transformation

综上,时域特征的表现形式  $R_f$  是长度为  $w$  的 SAX 单词,方法参数  $P_f$  包括子序列长度  $l$ 、SAX 单词长度  $w$  和离散化时的字符区间大小  $\Sigma$ 。时间序列  $T$  和特征  $f = \langle R_f, \langle l, w, \Sigma \rangle, L_f, \theta_f \rangle$  的相似性通过  $R_f$  和转换得到的所有 SAX 单词的最小距离计算:

$$D(T, R_f) = \min_{1 \leq i \leq |T| - l + 1} \text{dist}(\text{sax}(T_{i:l}), R_f) \quad (3)$$

其中,  $\text{sax}(T_{i:l})$  表示  $T$  的子序列通过参数  $P_f$  转换得到的结果,  $\text{dist}(\cdot, \cdot)$  计算两个 SAX 单词之间的距离。  $\theta_f$  和  $L_f$  的计算将在后文给出。

频域特征与时域特征类似,区别是时域特征将子序列转换为 SAX 单词,频域特征采用 SFA<sup>[9]</sup> 对子序列在频域进行转换。本文采用 SFA 转换方式,对长度为  $l$  的子序列应用离散傅里叶变换(Discrete Fourier Transform, DFT),将转换结果的前  $w$  个傅里叶系数根据多系数分桶(Multiple Coefficient Binning, MCB)的方法离散化,得到 SFA 单词。图 4 给出了对频域特征来源序列进行 SFA 转换的过程,以第一段长度  $l=100$  的子序列为例,最终转换为 bcdB 的 SFA 单词。

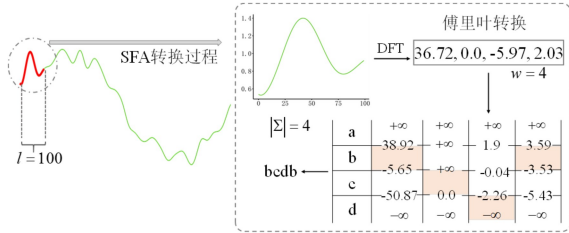


图 4 SFA 转换过程实例

Fig. 4 Example of SFA transformation

频域特征的表现形式  $R_f$  为 SFA 单词,转换参数  $P_f$  由子序列长度  $l$ 、单词长度  $w$  以及单词表  $\Sigma$  组成。给定时间序列  $T$ ,特征  $f$  和  $T$  的相似性度量方式和时域特征一致。尽管 SAX 单词和 SFA 单词在形式上相似,但 SAX 单词描述的是原始序列时域的形状,SFA 单词则描述时序频域上的形态。

时域特征和频域特征仅仅考虑了时间序列特定的形状特征,区间特征能够捕捉时间序列位于固定区间的信息,尤其适用于原始序列存在噪音的情况。本文选择 RISE<sup>[11]</sup> 中提供的两种转换方式,即自相关函数(Auto-correlation Function, ACF)和功率谱(Power Spectrum, PS),作为生成区间特征的方法。ACF 在一定的滞后范围内测量时间序列中子序列的相互依赖性;PS 是 ACF 的傅里叶变换,用于描述子序列之间的低频相关性。对于固定区间  $[s, e]$ ,  $s$  和  $e$  分别表示区间的起始点和终止点,ACF 和 PS 将区间转换成一个系数序列,从中选择最有区分力的系数。

因此,区间特征的转换参数表示为  $P_f = (s, e, \lambda, \tau)$ 。  $\lambda \in \{ACF, PS\}$  表示转换方式,  $\tau$  表示转换后的系数序列中选择哪一个系数,  $R_f$  表示系数序列中的典型值(typical value)。  $R_f$  和  $T$  之间的距离通过式(4)计算。

$$D(T, R_f) = P_f(T) - R_f \quad (4)$$

图 5 分别给出了时域特征和区间特征的示例。图 5(a)中,时域特征的表现形式  $R_f$  为 SAX 单词 AEF,时间序列  $T$  以参数  $P_f$  转换,得到一组 SAX 单词。如图所示,经距离计算,  $T_1$  是最接近特征  $f$  的时间序列。图 5(b)给出了区间

特征的案例,在区间  $[10, 40]$  上应用 ACF 转换,根据  $\tau=2$  选择需要所在列的系数值与  $R_f=5$  进行比较。

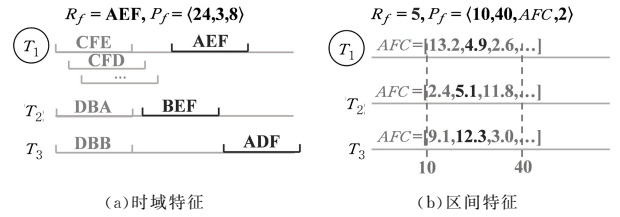


图 5 时域特征与区间特征示例

Fig. 5 Example of time-domain and interval-based feature

## 4.2 特征生成

本文针对单个维度的数据挖掘最有代表性的特征,首先需要为每个维度生成一组高质量的候选特征集合  $F_0$ 。HFC 使用类似决策树的树结构进行特征选择,这样的树称为特征树。本文使用  $K$  棵特征树并行提取特征。

### 4.2.1 特征树的构建

特征树的构建从根节点开始,迭代构建子树。对于一个节点,将其划分为  $|C|$  个子节点,  $|C|$  为当前节点数据集包含的类别个数。  $|C|=1$  时表示当前节点只存在同一类别的数据,当前分支划分结束。

特征树的构建与决策树的区别在于,决策树对节点的划分基于一个属性值,而特征树的划分基于  $|C|$  个同类型且由相同参数转换得到的特征。这  $|C|$  个特征称为节点特征组,每个特征的特征标签分别对应  $|C|$  个类别中的一类,即该特征仅用于判别时间序列是否属于其表示的特征标签类别。图 1 给出了由 6 个维度组成的 MTS 数据集,包含 4 个类别,因此构建特征树时,根节点被划分为 4 个分支。以特征树 1 为例,在根节点的划分过程中生成包含 4 个时域特征的节点特征组,其中每个特征代表一种类型,在椭圆框中显示。节点特征组将根节点划分为 4 个子节点,划分依据是数据与每个子节点对应的特征值的相似性。为了便于理解,图 1 给出了特征树的理想划分情况,即通过一次划分就将所有数据完美分到了对应的集合中,每个子节点包含一种类型的数据,划分结束。

对于每个节点的划分,本文为每种类型的特征生成  $N_g$  个节点候选特征组,分别计算每个候选特征组划分节点带来的信息增益,选择信息增益最大的候选特征组作为最终的节点特征组,用于划分节点数据,并将该特征组中的所有特征添加到候选特征集合  $F_0$  中。HFC 集成 3 种类型的特征,因此每个节点将从  $3 * N_g$  个节点特征组中选择。

### 4.2.2 节点特征的生成

上一小节从全局的角度介绍了整个特征树的构建过程,对于不同类型节点特征组的生成过程尚不明晰。本小节将分别介绍时域特征和区间特征的生成过程,频域特征与时域特征相似,只是使用的转换方法不同,下文不再赘述。

### 4.2.3 时域特征的生成

HFC 生成时域特征的节点特征组时,首先随机选取转换参数  $P_f = \langle l, w, \Sigma \rangle$ ,然后为每个类别  $c_i$  生成特征的表现形式,即 SAX 单词。具体来说,在该节点数据集中,随机选择

一个类别为  $c_i$  的样本时间序列  $T$ , 基于参数  $P_f$  对  $T$  进行转换, 得到一组 SAX 单词, 选择最有区分力的单词作为类别  $c_i$  的特征表现形式。

本文选择 TF-IDF 算法来评估符号单词对某个类别的重要程度, TF-IDF 值越大, 表示单词的区分力越强。单词  $\varphi$  的 TF-IDF 值可以通过计算词频 (Term Frequency, TF) 和逆文档频率 (Inverse Document Frequency, IDF) 的乘积得到。

TF 由式(5)计算得出, 其中  $F_{\varphi,c}$  表示单词  $\varphi$  在类别  $c$  的数据中出现的频数,  $N_c$  表示类别  $c$  的数据总数。

$$TF(\varphi, c) = \frac{F_{\varphi,c}}{N_c} \quad (5)$$

IDF 的计算公式为:

$$IDF(\varphi, c) = \log \frac{2}{I_{\varphi,c} + O_{\varphi,c}} \quad (6)$$

$I_{\varphi,c}$  和  $O_{\varphi,c}$  分别由式(7)和式(8)给出,  $F_{\varphi}$  表示单词  $\varphi$  在所有类别的数据中出现的频数。

$$I_{\varphi,c} = \begin{cases} 1, & F_{\varphi,c} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$O_{\varphi,c} = \begin{cases} 1, & F_{\varphi} - F_{\varphi,c} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

为每个类别生成时域特征的表现形式后, HFC 将节点数据集中除样本之外的时间序列通过  $P_f$  转换。对于提取的 SAX 单词, 使用式(3)计算它与其他时间序列的距离。对所有的距离值排序, 根据每个距离值划分节点数据集, 计算获得的信息增益, 信息增益最大的距离值为最佳分割点。至此, HFC 为类别  $c_i$  生成了一个完整的时域特征, 表现形式  $R_f$  为 SAX 单词, 转换参数为  $P_f$ , 代表类别为  $c_i$ , 阈值为最佳分隔点的数值。|C| 个时域特征共同构成一个节点特征组。

#### 4.2.4 区间特征的生成

对于区间特征, 算法首先随机选择一个区间  $[s, e]$  以及转换方式  $\lambda$  (ACF/PS), 对当前节点上所有时间序列位于该区间的子序列进行转换, 得到系数矩阵  $\mathbf{M}$ , 其中行代表时间序列, 列表代表系数。同样, HFC 为每个类别  $c_i$  生成特征。由于系数值为连续实数值, 因此算法选择信息增益作为评估选取哪个系数值作为区间特征的质量标准。

区间特征生成的具体过程是: 1) 对  $\mathbf{M}$  的每一列系数值排序, 计算每个系数值划分数据集带来的信息增益, 找到每列的最佳分割点; 2) 将最佳分割点对应信息增益最大的列作为区间特征选择的系数  $\tau$ , 作为转换参数的一部分, 该最佳分隔点即为特征的表现形式。注意, 对于不同的类别  $c_i$ , 相同系数值计算得到的信息增益不同。最后, HFC 得到类别  $c_i$  的区间特征, 参数  $P_f$  由区间的起始点  $s$  和  $e$ 、选取系数的所在列  $\tau$ 、转换方式  $\lambda$  组成, 最佳分隔点的数值为表现形式。

### 4.3 特征聚类

#### 4.3.1 低质量特征过滤

特征生成时的随机性在增强方法鲁棒性的同时, 也带来了一部分低质量特征, 且大量特征也不利于直接进行特征聚类。因此, 在特征聚类前, HFC 先将低质量特征从  $F_0$  中过滤, 获得一组高质量的特征组合, 记为  $F$ , 然后再进行特征聚类。

本文采用信息增益作为评估特征质量的标准。首先根据信息增益对  $F_0$  排序, 然后决定需要去除的低质量特征个数。筛选特征的过程依托于特征组合 (Feature Group, FG), 其定义如下。

**定义 8 (特征组合)** 给定排序的候选特征集合  $F_0$  和参数  $\delta$ , 特征组合  $FG_i = \{f_1, f_2, \dots, f_{i+\delta}\}$  是  $F_0$  的子集,  $FG_i \subset FG_{i+1}$ 。

因此, 筛除低质量特征的问题转变为寻找质量最高的特征组合的问题, 如果  $FG_i$  的质量最高, 则筛除其他不在  $FG_i$  中的特征。特征组合的质量由它对数据集  $D$  的分类结果准确度来衡量, 记为  $Q(FG_i)$ , 计算式如式(9)所示:

$$Q(FG_i) = \frac{|\{L(T) = \hat{L}(T) \mid T \in D\}|}{|D|} \quad (9)$$

其中,  $\hat{L}(T)$  代表特征组合对  $T$  的预测结果。

本文得到  $\hat{L}(T)$  的方式如下: 使用  $FG_i$  中的所有特征对时间序列  $T$  采用加权投票的方式预测  $T$  的所属类别, 将每个类别获得的票数权重之和作为该类别的最终得分, 选择得分最高的类别作为特征组合  $FG_i$  对  $T$  的预测类别  $\hat{L}(T)$ 。对于单个特征  $f$ , 若  $f$  识别  $T$  的类别为其代表类别  $L_f$ , 则  $f$  为  $T \in L_f$  投票, 权重为  $gain(f)$ ; 否则以  $-gain(f)$  作为权重投票。

HFC 计算每个特征组合  $FG_i$  的质量  $Q(FG_i)$ , 选择质量最高的  $FG_i$  作为最佳特征组合  $F$ , 继续下一阶段的特征聚类。

图 1 给出了  $\delta=2$  时, 低质量特征过滤的整个过程。假设排序后的候选特征集合为  $F_0 = \{f_1, f_8, f_6, f_4, \dots, f_3\}$ , 首先选择  $f_1$  和  $f_8$  作为初始特征组合, 根据式(9)计算该特征组合的质量  $Q(FG_1)$ , 图中简称为  $Q_1$ ; 接着, HFC 再选择 2 个特征  $f_6$  和  $f_4$  加入特征组合, 计算由 4 个特征组成的当前特征组合的质量  $Q(FG_2)$  ( $Q_2$ )。以此类推, 找到质量最高的特征组合, 作为最佳特征组合  $F$ , 进行特征聚类。

#### 4.3.2 多类型特征聚类

高质量特征组合  $F$  中存在相似和冗余特征, 本文设计聚类算法对这些特征进行聚合, 为每种类型的特征在每个类别下提取其代表性特征, 有效压缩特征空间, 提升准确度和结果的可解释性。如图 1 所示, 先将所有特征按照类型和代表类别划分到不同区域, 然后进行聚类。以时域特征的类别 1 为例, 最终将 5 个特征聚合得到两个代表性特征, 用实心图形标识。本文用  $RF_i (i \in [1, m])$  表示第  $i$  维度提取到的代表性特征 (Representative Feature), 所有维度提取到的代表性特征集合用  $RF$  表示。

不同类型的特征表征时间序列的不同特性, 时域特征和频域特征分别表示时域和频域有区分力的子序列, 区间特征表示固定区间存在的数据信息, 因此不同类型的特征聚类没有意义。不同类别的代表特征表示其类别蕴含的信息, 对不同类别的特征聚类同样没有意义。因此, 本文的聚类过程发生在同一类型、同一类别的特征集合中。首先将最佳特征组合  $F$  中的特征按照类型和代表类别划分为不同聚类区域; 对于区间特征, 按照 ACF 和 PS 两种转换方式进一步划分。图 6 通过一个二类特征聚类的示例解释上述特征区域的划分,

如图所示,最终聚类过程发生在7个特征集合中,将这样的特征集合称为同质特征集合(Homogeneous Feature Group, HFG)。

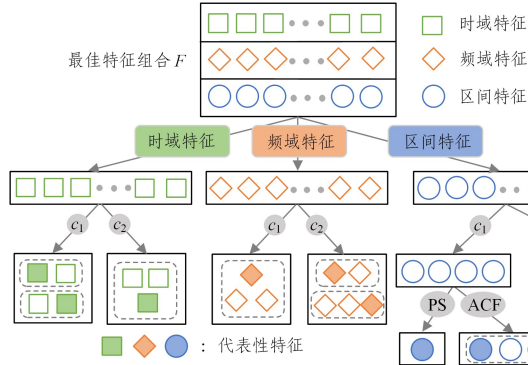


图6 多类型特征聚类示例

Fig. 6 Example of multi-type feature clustering

特征之间的距离定义是聚类的关键,本文为3种类型的特征分别提出了不同的距离计算方式,下文将依次具体介绍。

对于时域特征,由于SAX单词的长度不同,本文先将SAX单词逆变换回原始数值序列,再计算两个序列间的欧氏距离。给定一个HFG中的时域特征 $f$ 和 $f'$ ,计算式如式(10)所示:

$$d^{\text{time}}(f, f') = \min_{\epsilon \in [0, m-m']} \sqrt{\frac{\sum_{i=1}^m (t'(i) - t(i+\epsilon))^2}{m'}} \quad (10)$$

其中, $t=(t_1, t_2, \dots, t_m)$ 和 $t'=(t'_1, t'_2, \dots, t'_{m'})$ 表示单词逆变换得到的原始序列,长度分别为 $m$ 和 $m'$ , $m \geq m'$ 。

对于频域特征,SFA单词的每个字符独立表示离散傅里叶变换的基本函数。给定一个HFG中的频域特征 $f$ 和 $f'$ ,可以直接计算SFA单词之间的距离。但是,原始序列长度有明显差异的情况也需要考虑,式(11)使用长度比例作为乘积因子限制

$$d^{\text{freq}}(f, f') = \sum_{i=0}^l |R_f(i) - R_{f'}(i)| * \ln \frac{m}{m'} \quad (11)$$

其中, $l = \min(w, w')$ 。

对于区间特征,起始点 $s$ 和终止点 $e$ 分别用于定位区间所在位置,系数 $\tau$ 决定在转换后得到的系数序列中选择哪一个系数,因此,本文基于这些重要参数定义区间特征的距离。给定一个HFG中的区间特征 $f$ 和 $f'$ ,特征参数分别为 $P_f=(s, e, \tau)$ 和 $P_{f'}=(s', e', \tau')$ ,则其距离计算式如下:

$$d^{\text{int}}(f, f') = |s-s'| + |e-e'| + |\tau-\tau'| \quad (12)$$

确定每种类型特征的距离计算方式后,下文将讨论具体的聚类过程。最简单的策略是在每个HFG中独立聚类,然而这种策略可能无法获取全局最优解。因此,本文从全局视角对HFG进行聚合,所有HFG可以表示为一个序列 $HFG_i = \{HFG_1, HFG_2, \dots, HFG_{|H|}\}$ , $|H|$ 表示HFG的个数。以图6为例, $|H|=7$ 。

本文采用自顶向下的层次聚类作为聚类方法。初始状态下,每个HFG中的所有特征被视为一个簇。在聚类的过程中,每步选取同质特征集合HFG<sub>i</sub>中的一个簇,将其划分为两个子簇,使得划分后HFG<sub>i</sub>中所有子簇的特征组合在一起

得到的分类准确度相比划分前提升最大。

簇的划分采用贪心策略,两个子簇之间的距离定义为每个簇中两个点之间的最短距离。聚类结束的终止条件是:1)HFG<sub>i</sub>的分类准确度开始下降;2)每个HFG都被划分成5个子簇。聚类过程不断迭代,直到满足终止条件。

本文将每个HFG<sub>i</sub>划分成一些簇后,在每个簇中选取质量最高的特征作为代表性特征。在图6给出的示例中,每个HFG中深色的特征即为选择的代表性特征。特征质量由两个因素相乘获得:1)特征生成时的信息增益;2)特征的分类准确度,根据4.3.1节中的加权投票方式度量。代表性特征不仅考虑特征本身的重要性,同时需要表征所在簇的大小。特征所在的簇越大,特征代表的信息就越多。因此,代表性特征的投票权重设置为 $gain(f) * \ln(N_{\text{cls}})$ , $N_{\text{cls}}$ 表示特征所在簇cls中的特征总数。

#### 4.4 多维度特征转换

通过单维度的特征生成与聚类,本文为每个维度提取到少量的代表性特征,如何共同考虑多维度的特征成为关键问题。本节提出了多维度特征转换(MFT)方法,综合考虑各个维度特征,融合单维度不同类型的代表性特征。如图1所示,维度 $i$ 提取到代表性特征集合 $RF_i$ ,原始的MTS数据集由6个维度组成,因此提取到6个代表性特征集合,将所有特征对MTS的转换结果进行组合,得到最终用于分类的距离矩阵。下面给出MFT的定义。

**定义9(多维度特征转换)** 给定一组代表性特征集合 $RF = \{f_1, f_2, \dots, f_{|RF|}\}$ ,特征 $f_i$ 的来源维度由 $v_i (i \in |RF|)$ 表示。给定时间序列 $T$ ,通过公式 $d_i = \text{dist}(f_i, P_{f_i}(t_{v_i}))$ 计算特征 $f_i$ 与 $T$ 对应维度数据 $t_{v_i}$ 的距离,将 $T$ 转换为 $(d_1, d_2, \dots, d_{|RF|})$ 。

图7为多维度特征转换应用于BasicMotions数据集的示例。左侧数据源于BasicMotions中类别2的MTS实例,中间两个特征分别是第四维度和第六维度提取得到的代表性特征,用于识别类别2和类别4的数据。图中计算时间序列相应维度与特征的距离,将一个多维时间序列转换为右侧的距离值向量。

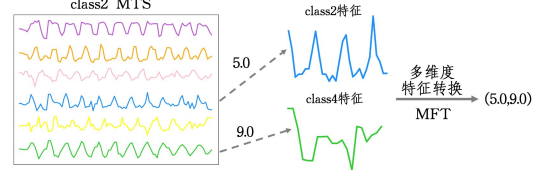


图7 多维度特征转换图示

Fig. 7 Illustration of multivariate feature transformation

对于MTS数据集 $D \in \mathbb{R}^{N \times m \times n}$ ,MFT将其转变为距离实值矩阵,大小为 $N \times |RF|$ 。该矩阵可以直接用于标准分类器学习分类模型。本文选择带线性核的SVM作为分类器。SVM的目标是找到一个超平面,将数据划分成两个类别,起初用于二分类问题。对于本文涉及的多分类问题,采用一对一法,即训练时在任意两个类别的数据之间设计一个SVM分类器,因此 $|C|$ 个类别的数据将设计 $|C| * (|C| - 1) / 2$ 个SVM分类器。SVM方法的训练过程和最后结果比较容易

理解,训练好的每个 SVM 分类器中,权重易于获得,有利于最终结果的可解释性。

## 5 实验结果与分析

### 5.1 实验设置

UEA 数据仓库<sup>[29]</sup>是评估多维时间序列分类方法的经典权威数据集,数据来源于现实生活中的不同应用场景,例如人类活动识别、ECG/EEG 信号分类等。MTS 数据集的维度范围为 2~1345,长度为 8~17984,训练集大小为 1~5890。本实验选取 20 个 MTS 数据集,去除了长度过长、维度过高的数据集,训练集与测试集的划分与原数据集保持一致。

实验环境为 64 位 Windows 10 单机环境,处理器为 Intel(R) Core(TM) i9-10900 CPU(2.81 GHz),内存 128 GB。

对于实验参数,本实验设置特征树的个数为  $k=100$ ,特征树每个节点生成不同类型特征的个数为  $N_s=100$ ,特征生成过程的最大树深为 10。对于时域特征和频域特征,滑动窗口长度的取值集合为  $\{10, \dots, n\}$ ,字符集大小  $\Sigma$  为 4,单词长度的取值集合为  $\{6, 8, 10, 12, 14, 16\}$ 。

### 5.2 比较方法

实验选择当前分类效果好、比较流行的方法进行比较。

1) 3 种 MTSC 基准方法<sup>[29]</sup>:实验选择 3 种经典的基准分类方法,分别是基于欧氏距离的  $ED_1$ 、基于维度独立 DTW 的  $DTW_1$  和基于维度依赖 DWT 的  $DTW_D$ 。

2) 基于 shapelet 的方法:ShapeNet<sup>[7]</sup>是目前最流行的 shapelet 方法,它将不同长度的 shapelet 嵌入到统一空间,选择一组最佳 shapelet,计算 shapelet 与 MTS 的距离,将结果作为分类模型的输入。

3) 基于字典的方法:WEASEL+MUSE<sup>[22]</sup>作为经典的 BOP(bag-of-patterns)模型,在所有维度中使用滑动窗口提取子序列中的 SFA 单词,将维度、窗口长度和 SFA 单词串联形成一个新的单词,得到每个单词的直方图。

4) 基于区间的方法:DrCIF<sup>[12]</sup>对 CIF 进行扩展,作为在 2021 年提出的 HIVE-COTE2.0 的组成部分。该方法使用随机选择区间上的 catch22 特征集、周期图、差分表示作为特征,实现基于区间的森林,是当前最先进的基于区间的方法之一。因此,本文选择 DrCIF 作为区间方法的代表。

5) 深度学习方法:MLSTM-FCN<sup>[25]</sup>是用于 MTSC 的通用深度学习框架,使用 LSTM 和 CNN 网络模拟时间序列中的顺序信息;TapNet<sup>[6]</sup>是近两年流行的 MTSC 模型,该方法使用递归神经网络和 CNN 捕捉时间序列的特征,训练注意力原型网络;ROCKET<sup>[30]</sup>利用随机卷积核转换时间序列,以训练线性分类器,被认为是准确度与效率都令人满意的分类方法<sup>[15]</sup>;MiniRocket<sup>[31]</sup>是 ROCKET 的优化,它使用一组更小且固定的卷积核,在基本保持准确度相同的条件下,比 ROCKET 快了 75 倍;NS<sup>[28]</sup>提出了一种用于时间序列的无监督表示学习方法,使用 SVM 作为最终的分类器。

6) 集成方法:XEM 是为多维时间序列分类设计的一种可解释框架,其基于一种新的混合集成方法,提供适用于单个 MTS 的解释。

### 5.3 准确性实验

表 1 列出了在 UEA 数据集的准确度比较结果,对比方法的准确度来源于原论文提供的数据或代码复现,以及通过 sktime 获取多维扩展方法,例如 ROCKET。图 8 给出了不同方法的临界差分图(Critical Difference Diagram),展示了各方法的分类排序差异。

表 1 UEA 多维时间序列数据集上的准确度比较

Table 1 Classification results comparison on UEA multivariate time series datasets

Dataset	MTSC 基准			区间	频域	时域	深度学习					集成	HFC
	$ED_1$	$DTW_1$	$DTW_D$	DC	WM	SN	MF	TN	RC	MRC	NS	XEM	
AWR	0.970	0.980	0.987	0.983	0.993	0.987	0.973	0.987	<b>0.997</b>	0.992	0.987	0.993	0.977
AF	0.267	0.267	0.200	0.333	0.253	0.400	0.267	0.333	0.067	0.133	0.133	<b>0.467</b>	<b>0.467</b>
BM	0.675	<b>1.000</b>	0.975	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.900	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
CT	0.964	0.969	0.990	0.980	0.985	0.980	0.985	<b>0.997</b>	0.992	0.993	0.994	0.979	0.987
CK	0.944	0.986	<b>1.000</b>	0.986	0.992	0.986	0.875	0.958	<b>1.000</b>	<b>1.000</b>	0.986	0.986	<b>1.000</b>
EP	0.667	0.978	0.964	0.998	<b>1.000</b>	0.987	0.768	0.971	0.986	<b>1.000</b>	0.957	0.986	<b>1.000</b>
EC	0.293	0.304	0.323	0.624	0.342	0.312	0.350	0.323	0.418	0.468	0.236	0.372	<b>0.639</b>
ER	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.200	0.267
FM	0.550	0.520	0.530	0.530	0.540	0.580	0.570	0.530	0.560	0.550	0.540	<b>0.590</b>	<b>0.590</b>
HMD	0.279	0.306	0.231	0.482	0.249	0.338	0.338	0.378	0.500	0.392	0.270	<b>0.649</b>	0.459
HW	0.371	0.509	<b>0.607</b>	0.358	0.286	0.451	0.267	0.357	0.593	0.507	0.533	0.287	0.419
HB	0.620	0.659	0.717	0.751	0.737	0.756	0.698	0.751	0.746	<b>0.771</b>	0.737	0.761	0.756
JV	0.924	0.959	0.949	0.979	0.922	0.984	<b>0.989</b>	0.965	0.968	<b>0.989</b>	<b>0.989</b>	0.978	0.962
LIB	0.833	0.894	0.872	0.907	0.921	0.856	0.856	0.850	0.900	0.922	0.867	0.772	<b>0.954</b>
NATO	0.850	0.850	0.883	0.863	0.920	0.883	0.772	0.939	0.872	0.928	<b>0.944</b>	0.916	0.911
PD	0.973	0.939	0.977	0.974	0.942	0.977	<b>0.986</b>	0.980	0.982	0.974	0.983	0.977	0.968
RS	0.868	0.842	0.803	0.897	0.878	0.882	0.750	0.868	0.895	0.868	0.862	0.941	<b>0.951</b>
SAD	0.967	0.960	0.963	0.982	0.980	0.975	0.986	0.983	<b>0.997</b>	0.993	0.956	0.973	0.982
SWJ	0.200	0.333	0.200	0.400	0.347	0.533	0.200	0.400	0.467	0.333	0.400	0.400	<b>0.600</b>
UW	0.881	0.869	0.903	0.917	0.906	0.906	0.875	0.894	<b>0.934</b>	N/A	0.884	0.897	0.906
Win/Ties	0	1	2	1	2	1	2	2	5	5	3	4	9

注:DrCIF;DC: WEASEL+MUSE;WM; ShapeNet;SN; MLSTM-FCN;MF; TapNet;TN; ROCKET;RC; MiniRocket;MRC。

由准确度的实验结果可以看出,本文方法 HFC 相比其他分类方法具有明显的优越性,在 11 个数据集上均取得了最佳结果。同时,由于 HFC 集成了不同类型特征,在 UEA

数据集上的分类结果稳定,不会出现分类准确度与最优准确度差距很大的情况。

在实验中的其他分类方法中,ROCKET 是最具竞争力的

分类方法,然而,ROCKET 通过随机卷积核分类,结果的可解释性低,不利于进一步挖掘分类结果的信息。ShapeNet 虽然通过提取 shapelet 获取结果的可解释性,然而其准确度在多个数据集上无法令人满意,且效率较差,在本文的实验环境下,大多数数据集耗时超过 5h。WEASEL+MUSE 采用 SFA 单词的直方图分类,HFC 同样采用 SFA 方法挖掘频域特征,且通过特征聚类压缩特征空间,使用少量代表性特征进行有效分类,因而 HFC 在多数数据集上的分类结果都优于 WEASEL+MUSE。DrCIF 在表 1 列出的分类结果的对比中处于劣势,但是在以区间特征为主的数据集上表现突出。EthanolConcentration(EC)<sup>[32]</sup>是由酒瓶中的水和乙醇溶液的原始光谱组成的数据集,浓度分为 35%,38%,40%,45% 4 类。通过实验测量得到的光谱组成,来确定任意瓶内样品的酒精浓度所属类别。由于波长的范围在紫外线、可见光、近红外线 3 个区间变化,每个区间携带的物理信息不同,因此,该数据集更适用于区间方法,使得 DrCIF 的分类准确度显著优于其他分类方法。HFC 同样集成区间方法,选取高质量的代表性特征,在 EC 数据集上的准确度仍然高于 DrCIF。

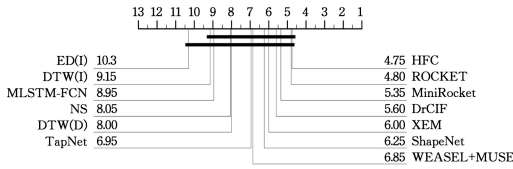


图 8 不同分类方法的临界差分图

Fig. 8 Critical difference diagram of different MTSC methods

#### 5.4 不同类型特征的影响

本文选择时域、频域、区间这 3 种类型特征构建分类模型,本节设计消融实验来评估不同类型特征对分类结果的影响。将 3 种类型中的任意 2 种特征相结合,得到 HFC 变体,保持参数设置一致,在 UEA 数据集上得到分类结果。图 9 给出了 3 种变体和 HFC 的比较结果,横轴表示变体的准确度,纵轴表示 HFC 准确度,对角线代表取得相同准确度,左上部分代表 HFC 比变体分类更准确,右下部分则相反。

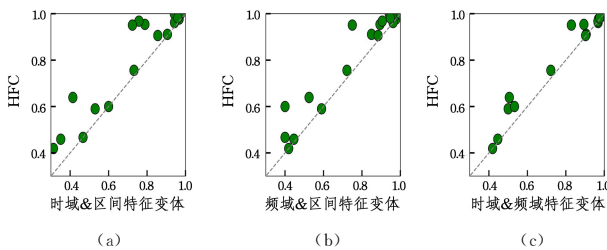


图 9 不同类型特征的重要性比较

Fig. 9 Contribution of different feature types

由图 9 可知,3 种变体的准确度都低于 HFC,证明了 3 类特征相结合的合理性与有效性。图 9(a)和图 9(b)位于左上方的点数多于图 9(c),表示时域特征和频域特征相较于区间特征对分类结果的贡献度更大。然而,一些数据集依赖于区间特征,例如 5.3 节提到的 EC 数据集,且区间特征对准确度的提升依然有一定贡献。为了使方法更好地泛化到更多类型的数据集上,保留区间特征依然是必要的。

#### 5.5 参数 K 的影响

本节实验研究树的个数  $K$  对准确率和效率的影响,图 10

给出了分类准确度随  $K$  的变化情况。图 10 显示, $K$  值小于等于 100 时,准确度随  $K$  的增大而上升,且越来越稳定;当  $K$  值大于 100 后,随着  $K$  的增大,在 200~500 时准确度出现下降的趋势;当  $K$  值大于 500 时,准确度均值不再有明显变化,且方差缩小,在各个数据集上的准确度趋于稳定。因此, $K$  值过大或过小都会导致准确度下降。 $K$  值太小,生成的候选特征数量少,削弱了算法筛选得到有区分力特征的可能性; $K$  值太大,虽然生成了大量特征,但为了保证分类结果的可解释性,本文规定每个 HFG 中的簇保持在 5 个以内,因此一些相似度低的特征会被聚在一个簇中,导致准确度降低。如果希望得到更准确的结果,需要增大聚类中簇的个数上限,但是这会使结果的可解释性下降。

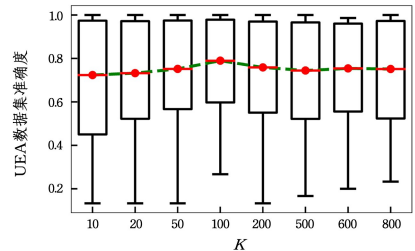


图 10 参数  $K$  对准确率的影响

Fig. 10 Influence of parameter  $K$  on accuracy

图 11 给出了不同 UEA 数据集上运行耗时随  $K$  的变化情况。Heartbeat(HB)和 SpokenArabicDigits(SAD)数据集分别代表 UEA 数据集中维度高和数量大的数据集,HB 由 61 个维度数据组成,SAD 包含 6599 个训练集,因此,当  $K$  小于 200 时,运行时间在这两个数据集上出现耗时曲线的拐点;当  $K \geq 500$  时,在大数据集上耗时增长速度加快。因此,本文选择  $K=100$  作为默认值来平衡高准确度、低方差和可接受的效率。

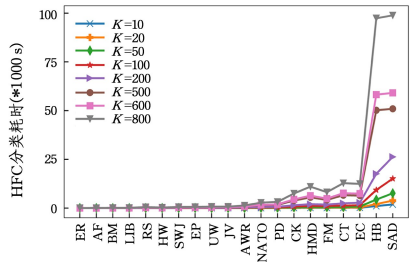


图 11 参数  $K$  对效率的影响

Fig. 11 Influence of parameter  $K$  on efficiency

#### 5.6 可解释性实验

本文提出的方法提取少量代表性特征,且易于可视化,使得分类结果具有可解释性。本实验从 UEA 数据仓库中选取 3 个数据集,即 Epilepsy, BasicMotions 和 EthanolConcentration,分别展示频域特征、时域特征和区间特征在分类结果中的可解释性,将 SVM 分类器训练得到的权重高的特征进行可视化分析。

Epilepsy 数据集<sup>[33]</sup>收集参与者模拟走路、跑步、锯切、癫痫发作 4 项不同类型的活动时,手腕上所佩戴的三轴加速度计记录的数据,包括 3 个维度、4 个类别。图 12 给出了 Epilepsy 数据集提取频域特征的案例。

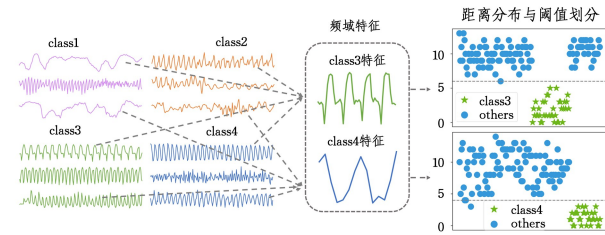


图 12 Epilepsy 数据集可解释性案例

Fig. 12 Interpretable case of Epilepsy

可视化使得分类方法对数据的划分过程变得清晰,图 12 左侧显示了 4 个类别的 MTS 实例,每个实例包含 3 个维度;中间虚线框图中将第一维度和第三维度提取到的两个频域特征可视化,代表类别分别为类别 3 和类别 4,可视化呈现出的特征形态明确了算法如何区分对应维度的不同类别;右侧呈现了特征与所在维度数据的距离分布,以及对距离的阈值划分。可以看到,阈值清晰地划分了不同类别的距离分布,例如,代表类别为 3 的特征根据距离阈值将类别 3 的数据和其他类别的数据区分开。因此,该实验证明了 HFC 对于多维时间序列能够准确选择有区分力的特征以及 MFT 表示的合理性。

BasicMotions 数据集由参与者佩戴智能手表进行 4 项活动时记录产生,因此数据集包含 4 个类别,分别是行走、休息、跑步和打羽毛球,根据传感器的位置不同,每个数据由 6 个维度组成。图 13 给出了 HFC 应用于该数据集提取时域特征的案例。HFC 提取来源于维度 1 的代表类别 2 的时域特征和来源于维度 4 的代表类别 4 的时域特征,可以看到,每个特征都反映了其所在类别数据存在的特别形态。根据这两个特征,可以将特征的代表类别数据和其他类别数据通过距离阈值明显区分开来。

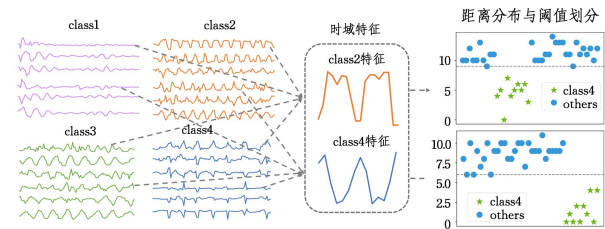


图 13 BasicMotions 数据集可解释性案例

Fig. 13 Interpretable case of BasicMotions

EthanolConcentration 数据集<sup>[32]</sup>在 5.3 节有提及,其包括 3 个维度、4 个类别。该数据集的分类标准主要与不同区间包含的信息相关,且酒瓶形状、厚度的不同,以及标签等障碍物的存在,造成数据存在较多噪音,使得该数据集适用于区间特征。

图 14 给出了来源于维度、代表类别为 4 的区间特征对数据集的划分实例。图中左侧为第一维度 4 种类型的数据,两条竖直虚线之间表示特征找到的区间,右侧给出根据数据和代表类别为 4 的特征的距离分布图。尽管中间区间(可见光波长区间)存在大幅变化,区间特征仍然能准确寻找到包含对乙醇浓度最有鉴别力信息的近红外区间。在距离分布图中可以看到阈值将大部分的数据正确划分,由于 EthanolConcentration 数据较难划分,因此无法通过一个特征将所有数据

划分正确,存在一些识别错误的数据是正常的。

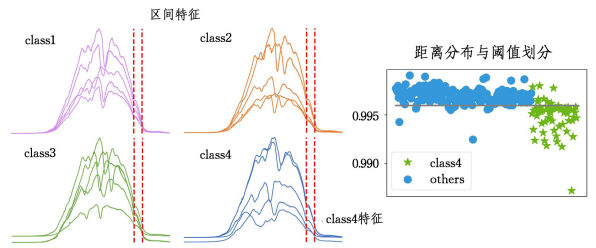


图 14 EthanolConcentration 数据集可解释性案例

Fig. 14 Interpretable case of EthanolConcentration

**结束语** 本文提出了一种基于异构特征融合的多维时间序列分类算法 HFC,以先分后合的方式,先提取不同维度的多类型代表性特征,再通过多维度特征转换融合各维度不同类型的特征空间。最终,多维时间序列转换为所有特征和数据的距离矩阵,作为标准分类器 SVM 的输入,训练分类模型。对于每个维度的特征提取,算法基于树结构生成时域、频域、区间这 3 种类型的候选特征,在此基础上提出聚类算法,筛选低质量、相似和冗余的特征,得到少量代表性特征,极大地压缩了特征空间,提高了分类准确度,增强了结果的可解释性。在公开数据仓库 UEA 上的实验结果证明了 HFC 相比现有方法在准确度和解释性上的优越性。然而,UEA 仓库中提供的数据是处理过的数据,因此 HFC 可以直接应用,尚未考虑缺失值、数据流等情况。未来将在更多真实场景的数据集上应用本文方法,以应对更多的挑战。除此之外,本文对代表性特征的可解释性需要使用者根据训练权重自行使用和展示,可解释性作为 HFC 特有的优点,可用性还需要进一步提升,未来考虑进一步评估特征的可解释性,将特征与分类过程的展示可视化。

### 参 考 文 献

- [1] SEZER O B, GUDELEK M U, OZBAYOGLU A M. Financial time series forecasting with deep learning: A systematic literature review: 2005—2019[J]. Applied Soft Computing, 2020, 90: 106181.
- [2] QI H, XIAO S, SHI R, et al. COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis[J]. Science of the Total Environment, 2020, 728: 138778.
- [3] RUBWURM M, KÖRNER M. Self-attention for raw optical satellite time series classification[J]. ISPRS Journal Of Photogrammetry and Remote Sensing, 2020, 169: 421-435.
- [4] SILVA D F, GIUSTI R, KEOGH E, et al. Speeding up similarity search under dynamic time warping by pruning unpromising alignments[J]. Data Mining and Knowledge Discovery, 2018, 32(4): 988-1016.
- [5] YE L, KEOGH E. 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.
- [6] LE NGUYEN T, GSPONER S, IFRIM G. Time series classification by sequence learning in all-subsequence space[C]// 2017 IEEE 33rd International Conference on Data Engineering (ICDE). IEEE, 2017: 947-958.

- [7] LI G, CHOI B, XU J, et al. Shapenet: A shapelet-neural network approach for multivariate time series classification[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(9):8375-8383.
- [8] LIN J, KEOGH E, WEI L, et al. Experiencing SAX: a novel symbolic representation of time series[J]. Data Mining and Knowledge Discovery, 2007, 15(2):107-144.
- [9] SCHÄFER P, HÖGQVIST M. SFA: a symbolic fourier approximation and index for similarity search in high dimensional datasets[C] // Proceedings of the 15th International Conference on Extending Database Technology. 2012:516-527.
- [10] DENG H, RUNGER G, TUV E, et al. A time series forest for classification and feature extraction[J]. Information Sciences, 2013, 239:142-153.
- [11] LINES J, TAYLOR S, BAGNALL A. Hive-cote: The hierarchical vote collective of transformation-based ensembles for time series classification[C] // 2016 IEEE 16th International Conference on Data Mining (ICDM). IEEE, 2016:1041-1046.
- [12] MIDDLEHURST M, LARGE J, FLYNN M, et al. HIVE-COTE 2.0: a new meta ensemble for time series classification[J]. Machine Learning, 2021, 110(11):3211-3243.
- [13] SHIFAZ A, PELLETIER C, PETITJEAN F, et al. TS-CHIEF: a scalable and accurate forest algorithm for time series classification[J]. Data Mining and Knowledge Discovery, 2020, 34(3):742-775.
- [14] ISMAIL FAWAZ 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.
- [15] RUIZ A P, FLYNN M, LARGE J, et al. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances[J]. Data Mining and Knowledge Discovery, 2021, 35(2):401-449.
- [16] ZHANG X, GAO Y, LIN J, et al. Tapnet: Multivariate time series classification with attentional prototypical network[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(4):6845-6852.
- [17] KARLSSON I, PAPAPETROU P, BOSTRÖM H. Generalized random shapelet forests[J]. Data Mining and Knowledge Discovery, 2016, 30(5):1053-1085.
- [18] 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 2015 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics. 2015:289-297.
- [19] WISTUBA M, GRABOCKA J, SCHMIDT-THIEME L. Ultrafast shapelets for time series classification[J]. arXiv:1503.05018, 2015.
- [20] BAYDOGAN M G, RUNGER G. Time series representation and similarity based on local autopatterns[J]. Data Mining and Knowledge Discovery, 2016, 30:476-509.
- [21] BAYDOGAN M G, RUNGER G. Learning a symbolic representation for multivariate time series classification[J]. Data Mining and Knowledge Discovery, 2015, 29:400-422.
- [22] SCHÄFER P, LESER U. Multivariate time series classification with WEASEL+ MUSE[J]. arXiv:1711.11343, 2017.
- [23] MIDDLEHURST M, LARGE J, BAGNALL A. The canonical interval forest (CIF) classifier for time series classification[C] // 2020 IEEE International Conference on Big Data. IEEE, 2020:188-195.
- [24] LUBBA C H, SETHI S S, KNAUTE P, et al. catch22: Canonical time-series characteristics[J]. Data Mining and Knowledge Discovery, 2019, 33(6):1821-1852.
- [25] KARIM F, MAJUMDAR S, DARABI H, et al. Multivariate LSTM-FCNs for time series classification[J]. Neural Networks, 2019, 116:237-245.
- [26] ZHENG Y, LIU Q, CHEN E, et al. Time series classification using multi-channels deep convolutional neural networks[C] // International conference on web-age information management. Springer International Publishing, 2014:298-310.
- [27] TUNCEL K S, BAYDOGAN M G. Autoregressive forests for multivariate time series modeling[J]. Pattern Recognition, 2018, 73:202-215.
- [28] FRANCESCHI J Y, DIEULEVEUT A, JAGGI M. Unsupervised scalable representation learning for multivariate time series[J]. arXiv:1901.10738, 2019.
- [29] BAGNALL A, DAU H A, LINES J, et al. The UEA multivariate time series classification archive, 2018[J]. arXiv:1811.00075, 2018.
- [30] DEMPSTER A, PETITJEAN F, WEBB G I. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels[J]. Data Mining and Knowledge Discovery, 2020, 34(5):1454-1495.
- [31] DEMPSTER A, SCHMIDT D F, WEBB G I. Minirocket: A very fast (almost) deterministic transform for time series classification[C] // Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021:248-257.
- [32] LARGE J, KEMSLEY E K, WELLNER N, et al. Detecting forged alcohol non-invasively through vibrational spectroscopy and machine learning[C] // Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham: Springer, 2018:298-309.
- [33] VILLAR J R, VERGARA P, MENÉNDEZ M, et al. Generalized models for the classification of abnormal movements in daily life and its applicability to epilepsy convulsion recognition[J]. International Journal of Neural Systems, 2016, 26(6):1650037.



**QIAO Fan**, born in 1998, postgraduate. Her main research interests include database, data mining, and information retrieval.



**WANG Peng**, born in 1979, Ph.D. professor, is a member of CCF (No. 41708M). His main research interests include database, data mining, and series data processing.