



计算机科学

COMPUTER SCIENCE

基于优化和两阶段筛选的时间序列Shapelets提取研究

李晨, 万源

引用本文

李晨, 万源. 基于优化和两阶段筛选的时间序列Shapelets提取研究[J]. 计算机科学, 2023, 50(2): 146-157.

LI Chen, WAN Yuan. [Study on Time Series Shapelets Extraction Based on Optimization and Two-phase Filtering](#) [J]. Computer Science, 2023, 50(2): 146-157.

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)

[基于点云数据的交通环境下单阶段三维目标检测方法](#)

Single-stage 3D Object Detector in Traffic Environment Based on Point Cloud Data
计算机科学, 2022, 49(11A): 210900079-6. <https://doi.org/10.11896/jsjcx.210900079>

[基于注意力机制与混合监督学习的钢轨表面缺陷检测模型](#)

Rail Surface Defect Detection Model Based on Attention Module and Hybrid-supervised Learning
计算机科学, 2022, 49(11A): 210800241-6. <https://doi.org/10.11896/jsjcx.210800241>

[基于相似度矩阵学习和矩阵校正的无监督多视角特征选择](#)

Unsupervised Multi-view Feature Selection Based on Similarity Matrix Learning and Matrix Alignment
计算机科学, 2022, 49(8): 86-96. <https://doi.org/10.11896/jsjcx.210700124>

[基于贝叶斯攻击图的动态网络安全分析](#)

Dynamic Network Security Analysis Based on Bayesian Attack Graphs
计算机科学, 2022, 49(3): 62-69. <https://doi.org/10.11896/jsjcx.210800107>

[基于YOLO优化的轻量级目标检测网络](#)

Light-weight Object Detection Network Optimized Based on YOLO Family
计算机科学, 2021, 48(11A): 265-269. <https://doi.org/10.11896/jsjcx.201000152>

基于优化和两阶段筛选的时间序列 Shapelets 提取研究

李 晨 万 源

武汉理工大学理学院 武汉 430070

(lichenstatistics@whut.edu.cn)

摘 要 与基于全局特征的时间序列分类方法相比,基于 shapelets 的分类方法在可解释性和分类速度方面更具优势。针对现有的优化模型学习到的 shapelets 判别力不足以及 shapelets 候选数量太多等问题,提出了基于优化和两阶段筛选的时间序列 shapelets 提取算法。首先对时间序列取样,结合极值点和序列趋势对取样的时间序列进行分组,根据分组结果对稀疏组 Lasso 正则器的每项赋予权重,并在加权稀疏组 Lasso 的每一组中都使用融合罚正则项来保证解的相邻位置平坦变化,将多项稀疏正则项作为正则器与局部线性判别分析相结合来构建目标函数。然后,建立一个两阶段的筛选框架来度量组的稀疏性,从而快速找到对分类起决定性作用的关键组。最后仅使用一组关键组来提取 shapelets 用于时间序列的分类,缩小了 shapelets 的规模。在 28 个时间序列数据集上进行了大量实验,实验结果表明,与现有的基于 shapelets 的提取方法相比,所提方法不仅能显著提高分类精度,具有较高的时间效率,而且能够在一定程度上缩小 shapelets 的规模。

关键词: shapelets; 两阶段筛选框架; 加权稀疏组 Lasso; 融合罚; 关键组

中图分类号 TP391

Study on Time Series Shapelets Extraction Based on Optimization and Two-phase Filtering

LI Chen and WAN Yuan

School of Science, Wuhan University of Technology, Wuhan 430070, China

Abstract Compared with the time series classification methods based on global features, the shapelet-based methods have more advantages in interpretability, efficiency and accuracy. In order to solve the problems of insufficient discrimination of shapelets obtained from existing sparse models and the large scale of shapelets candidates, this paper proposes a shapelets extraction method based on optimization and two-phase filtering. First, the time series are sampled, and the sampled time series are grouped by combining the extreme points and the trend, then the weight of each item in the sparse group lasso regularizer are assigned according to the grouping results. The fused penalty regularization is used in each group of the weighted sparse group lasso to ensure that the adjacent positions of the solution change smoothly. Those sparse regularization terms are combined as constraints to construct the objective function together with the local fisher discriminant analysis. Then, a two-phase filtering framework is established to measure the sparsity of groups, so as to quickly locate the key group that plays a decisive role in classification. Finally, this key group is retained to extract shapelets for time series classification, which reduces the candidates of shapelets. Extensive experiments are carried out on 28 datasets, and the experimental results show that, compared with the existing shapelets-based extraction methods, the proposed method significantly improves the classification accuracy with a good efficiency, and reduces the scale of shapelets to a certain extent.

Keywords Shapelets, Two-phase filtering framework, Weighted sparse group Lasso, Fused penalty, Key group

1 引言

时间序列分类在金融^[1]、医疗卫生^[2]、灾害监测^[3]、生物信息学^[4]等领域都有广泛的应用。传统的时间序列分类方法大多利用全局特征,即使用整条时间序列作为特征来进行分类。相似性度量是时间序列分类的重要依据^[5],主要使用欧氏距离和动态时间弯曲距离等来衡量。例如使用动态时间

弯曲(Dynamic Time Warp, DTW)^[6]度量距离的动态时间弯曲最近邻(DTW-1NN)^[7]等传统方法存在时间和存储空间双重消耗、可解释性差等问题。因此,如何提取有效的特征并提升分类的性能和可解释性是时间序列分类的重要研究问题。

为解决存储成本和时间成本高及可解释性差等问题, Ye 等^[8]于 2009 年提出了 shapelets 的概念。Shapelets 是从所有子序列中挑选出来的具有类别辨别能力的子序列,基于

到稿日期:2021-12-06 返修日期:2022-04-22

基金项目:中央高校基本科研业务费专项资金(2021III030JC)

This work was supported by the Fundamental Research Funds for the Central Universities of Ministry of Education of China(2021III030JC).

通信作者:万源(wanyuan@whut.edu.cn)

shapelets 的分类方法在分类精度、分类效率、鲁棒性和可解释性上通常优于传统的时间序列分类方法。

shapelets 的提取方法大致可分为基于搜索的提取方法和基于学习的提取方法。

基于搜索的提取方法通常用信息增益^[8]、Kruskal-Wallis 检验^[9]、Mood 中位数检验^[10]和 F 统计量^[11]等度量 shapelets 的质量,以筛选出具有最大判别力的 shapelets。Ye 等用暴力算法^[8]寻找具有判别力的 shapelets,需存储所有的候选 shapelets,时间和空间消耗都很大。之后他们提出子序列早期放弃和早期熵剪枝^[8]两种策略来加速 shapelets 的搜索,但不适用于大型数据集。Mueen 等^[12]使用高效的距离计算并提出新的熵剪枝,其缺点是难以在时间和搜索空间之间找到平衡。符号聚合近似(Symbolic Aggregate approximation, SAX)是由 Lin 等^[13]提出的将原始时间序列转换为离散和低维表示的另一种加速方法,高效的符号序列相似性度量技术可提高计算效率,该方法得到了广泛的应用^[14-15],但 SAX 单词的长度限制了分类的精度。Buhler 等^[16]引入随机投影寻找 shapelets, Rakthanmanon 等^[17]将该方法应用到 SAX 中,进一步加速了 shapelets 的搜索。Kruskal^[9]通过定义参数来限制候选 shapelet 的长度范围,减小了 shapelets 候选集的规模。还有一些方法使用一些随机的 shapelets^[18-19],但需要大量的 shapelets 来保证高精度,减慢了预测的速度,而且随机抽样使得所选择的 shapelets 的可解释性较差。Li 等^[20]先对原始的时间序列进行主成分提取,再对降维后的序列进行 shapelet 提取,但部分重要信息的丢失使得精度有所降低。

这些加速方法在一定程度上减少了 shapelets 候选集的数量,但获得的 shapelets 候选基数仍然很大,加速效果仍不理想。另一种通过优化目标函数来学习 shapelets 的位置的方法,为 shapelets 的提取提供了新的思路。该方法被称为基于学习的 shapelets 提取方法,所学习到的 shapelets 可以是来自原始序列的子序列,也可以是变换得到的新序列。

最早使用基于学习的方法来获取 shapelets 的是 Grabocka 等^[21]提出的学习时间序列 shapelets(Learning Time-series Shapelets, LTS)。他们采用特定的随机梯度下降法求解逻辑损失函数,避免了尝试大量的候选对象,可直接学习近似最优的 shapelets。LTS 的缺点是耗时较长,不适用于大型数据集。Hou 等^[22]提出了融合 Lasso 的广义特征向量方法(Fused Lasso Generalized eigenvector method, FLAG),该方法利用广义特征向量法^[23]寻找投影的优势并融合 Lasso^[24]的稀疏性来学习 shapelets,运行时间较 LTS 大大缩短,但其分类准确率也有所下降。Zhang 等^[25]提出了一种基于维数选择的 shapelets 提取方法(Shapelets Extraction by Dimensionality Selection, SEDS),该方法利用局部线性判别分析^[26]和融合 Lasso 正则器直接学习 shapelets 的位置。与 LTS 相比,SEDS 分类的准确性和效率均有所提高,但是其分类速率较 FLAG 有所下降。Wang 等^[27]利用卷积神经网络^[28]来处理时间序列分类任务,引入一个对抗性正则化的约束模型来学习更多具有可解释性的 shapelets,在精度和可解释性上都提高,但也存在运行时间相对较长的缺点。Zhao 等^[29]

使用最小二乘函数和融合 Lasso 正则器构建多元时间序列的 shapelets 学习模型,先对每维一元序列进行分类,再投票决定多元时间序列的分类结果,但其忽略了多元时间序列中每维一元序列之间的相关性。

上述以学习的方法提取 shapelets 的模型有些能够一定程度地提升运行效率,但判别力不足;有些模型学习到的 shapelets 候选数量仍较多,模型可解释性不足。针对这些问题,本文提出一种基于优化和两阶段筛选的时间序列 shapelets 提取研究算法(the Shapelets Extraction based on optimization and 2-Phase Filtering, SE2PF)。本文的创新点为:

(1)针对现有模型学习到的 shapelets 判别力不足的问题,首次将稀疏组 Lasso 引入 shapelets 的学习中,根据分组结果赋予权重,构成加权稀疏组 Lasso 正则项,并在此基础上引入融合 Lasso 的融合罚正则项,以此构成正则器,保证所构建的稀疏模型的合理性。

(2)针对稀疏组 Lasso 模型的预设组效应特点,提出结合采样的时间序列的极值点位置与趋势的分组方法。

(3)针对学习到的 shapelets 候选数量仍较多及模型可解释性不足的问题,建立两阶段筛选框架来寻找关键组,用关键组提取 shapelets,有效缩减了 shapelets 候选的规模,也使得获得的具有判别力的 shapelets 的位置更加明确清楚,提高了模型的可解释性。

(4)将构建的 shapelets 学习模型应用于 28 个覆盖多领域的数据集,实验结果证明,SE2PF 在 15 个数据集上获得了最高的分类准确率,在其他数据集上的准确率的排名也靠前。在运行时间上,SE2PF 能够在不显著降低分类速率的同时提高分类精度。

2 相关工作

2.1 融合 Lasso 正则器

基于学习方式获取 shapelets 的算法一般可总结为一个优化函数,优化函数的正则项在 shapelets 位置的学习中占有重要地位。融合 Lasso^[24]是近年来应用广泛的稀疏学习模型,其正则项如下:

$$\alpha \|\phi\|_1 + \beta \sum_{i=2}^l \|\phi_i - \phi_{i-1}\|_1, \phi \in \mathbb{R}^{n \times 1}$$

其中,第一项实现解的稀疏性,第二项融合罚使得解向量相邻位置的值尽可能接近,平坦变化,保留了变量的先后顺序。上述正则项也可改写为:

$$\alpha \|\phi\|_1 + \beta \|\mathbf{B}\phi\|_1, \phi \in \mathbb{R}^{n \times 1} \quad (1)$$

其中, $\mathbf{B} \in \mathbb{R}^{(n-1) \times n}$ 是一个稀疏矩阵,定义如下:

$$\mathbf{B}_{ij} = \begin{cases} 1, & j = i+1, i=1, 2, \dots, n-1 \\ -1, & j = i, i=1, 2, \dots, n-1 \\ 0, & \text{其他} \end{cases} \quad (2)$$

2.2 FLAG 与 SEDS

FLAG^[22]将广义特征向量法与融合 Lasso 的正则项相结合,构造了一个能直接学习时间序列 shapelets 位置的稀疏模型,其目标函数为:

$$\min_v v^T \mathbf{C}_j v + \alpha_1 \|\mathbf{B}v\|_1 + \alpha_2 \|v\|_1$$

$$\text{s. t. } \mathbf{v}^T \mathbf{C}_j \mathbf{v} = 1$$

其中, \mathbf{C}_j 是第 j 类的样本协方差矩阵。广义特征向量法用于寻找使第 i 类和第 j 类之间的投影方差比率最大的方向, 当一个类被压缩时, 最大程度地保留另一个类的信息。融合 Lasso 的正则项用来保证获得的解是块稀疏的, 同时保证时间的先后顺序信息。

SEDS^[25] 将局部线性判别分析 LFDA 与融合 Lasso 的正则项相结合来构造直接学习 shapelets 位置的稀疏模型, 融合 Lasso 的正则项的作用和 FLAG 相同。SEDS 的目标函数为:

$$\min_{\mathbf{v}} \mathbf{v}^T \mathbf{S}_w \mathbf{v} + \alpha_1 \|\mathbf{B}\mathbf{v}\|_1 + \alpha_2 \|\mathbf{v}\|_1$$

$$\text{s. t. } \mathbf{v}^T \mathbf{S}_b \mathbf{v} = 1$$

其中, \mathbf{S}_w 和 \mathbf{S}_b 分别表示局部线性判别分析的局部类内离散度矩阵和局部类间离散度矩阵。LFDA 保证投影后的类间距离大, 类内距离小, 同时保持数据的局部邻域结构。

多元时间序列的 shapelets 学习方法 SL_{MTSC}^[29] 将最小二乘函数和融合 Lasso 正则项相结合, 其目标函数如下:

$$\min_{\mathbf{v}} \|\mathbf{X}\mathbf{v} - \mathbf{y}\|_2^2 + \alpha_1 \|\mathbf{B}\mathbf{v}\|_1 + \alpha_2 \|\mathbf{v}\|_1$$

该方法将多元时间序列拆分为多个一元序列的分类问题, 投票决定多元序列的分类结果。

2.3 稀疏组 Lasso 正则器

稀疏组 Lasso (Spares Group Lasso, SGL)^[30] 是将 Lasso^[31] 与组 Lasso^[32] 的罚函数相结合的正则化稀疏模型。Lasso 实现变量的稀疏性, 组 Lasso 实现变量组的稀疏性, SGL 可同时实现图 1 所示的两种稀疏效应, SGL 的正则项如下:

$$\alpha \|\boldsymbol{\phi}\|_{2,1} + \beta \|\boldsymbol{\phi}\|_1, \boldsymbol{\phi} \in \mathbb{R}^{n \times 1} \quad (3)$$

其中, $\alpha \|\boldsymbol{\phi}\|_{2,1}$ 实现变量组的组稀疏性, $\beta \|\boldsymbol{\phi}\|_1$ 实现分组内的变量的稀疏性。

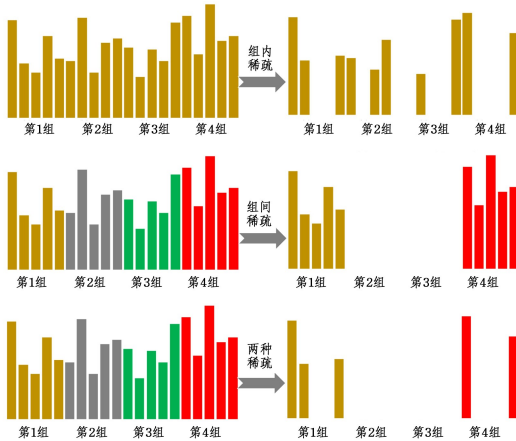


图 1 稀疏组 Lasso 的双重稀疏效应图

Fig. 1 Dual sparsity effect graph of sparse group Lasso

SGL 因预设组效应需要预先设定分组。通常, 同一个组的变量的回归系数同时为零时, 从稀疏模型中移除该组; 同时为非零时, 选中这些非零的组, 从组中实现变量的稀疏性。

2.4 符号描述

表 1 列出了文中使用的符号, 后续使用时将进一步解释每个符号的具体含义。在本文中, 使用大写黑体字母例如 \mathbf{X} 表示矩阵, 小写黑体字母例如 \mathbf{y} 表示向量。

表 1 本文涉及的符号

符号	描述
l	时间序列长度
J	每个数据集的组数
$m_j, j=1, 2, \dots, J$	每组序列的尺寸
$\mathbf{v}^{(j)} \in \mathbb{R}^{m_j \times 1}$	每组时间序列的投影向量
$\mathbf{A}_j \in \mathbb{R}^{m_j \times l}$	每组的系数矩阵
\mathbf{B}_j	每组的稀疏矩阵
\mathbf{S}_w	局部类内离散度矩阵
\mathbf{S}_b	局部类间离散度矩阵
α, β, γ	稀疏约束参数
ρ_1, ρ_2, ρ_3	惩罚参数
T	最大迭代次数
γ	停止阈值
$\mathbf{f}, \mathbf{a}_j, \mathbf{e}_j$	辅助向量
$\zeta_j, \boldsymbol{\mu}, \boldsymbol{\sigma}_j, \lambda$	拉格朗日乘子

3 本文方法

为了提取判别性更好的 shapelets, 同时缩小 shapelets 的规模, 提高 shapelets 学习模型的解释性, 本文提出了基于优化和两阶段筛选的时间序列 shapelets 提取研究 (SE2PF) 算法。本文引入 SGL, 结合模型特点先对采样的时间序列进行分组。根据分组结果对每组赋予权重, 使用加权的 SGL 正则项与融合 Lasso 的融合罚正则项来构造正则器, 正则器与局部线性判别分析相结合构造优化函数以获得每一组的投影向量, 再提出两阶段筛选框架来评估每一组投影向量的稀疏度, 加速关键组的定位。经两阶段的过滤, 仅剩两组作为关键组候选。最后分别使用两组候选单独提取 shapelets 以及训练 SVM, 对测试集分类精度最高的一组为关键组。因关键组的位置明确以及长度小于整条时间序列, 最终用于分类的 shapelets 仅仅来自关键组所在区域, 既缩小了 shapelets 候选集的规模, 又增强了 shapelets 学习模型的可解释性。

SE2PF 的框架图如图 2 所示。

3.1 分组

FLAG 与 SEDS 等学习模型是从全长时间序列提取 shapelets, 学习到的 shapelets 位于时间序列的各个位置, shapelets 位置的可解释性较弱。结合稀疏组 Lasso 模型的预设组效应及其可以选择少数组的原理, 先对时间序列分组。

3.1.1 时间序列采样

时间序列训练集中每个序列都有一个确定的类。通过时间序列的采样^[33-34] 可大大减少序列的数量。受此启发, 本文通过采样标准时间序列来完成后续分组。假设每一类中有 n_c 条时间序列, 时间序列的长度为 l , 针对每一类时间序列 $D_c = \{T_1, T_2, \dots, T_i, \dots, T_{n_c}\}$, 以及该类中的每一条序列 $T_i = \{t_1, t_2, \dots, t_j, \dots, t_l\}$, 通过下式选择最接近平均值的时间序列作为标准时间序列:

$$\text{sum}_i = \sum_{j=1}^a t_j, \text{mean} = \frac{\sum_{i=1}^{n_c} \text{sum}_i}{n_c} \quad (4)$$

$$T_c = \arg \min_{i \in \{1, 2, \dots, n_c\}} (|\text{sum}_i - \text{mean}|) \quad (5)$$

根据式 (4) 与式 (5) 寻找时间序列训练集中每一类对应的标准时间序列, 对训练数据集进行采样。

3.1.2 确定关键点

采样时间序列之后需要确定分组的关键词, 先寻找候选

关键点,再从中确定关键点的位置。标准时间序列中所有的极值点均为关键点候选。

关键点的确定分为初步选择以及后续调整。如果某个时间戳在采样的时间序列中都是极值点对应的公共时间戳,则可考虑选取该时间戳为分组的关键点位置;如果不能找到这样的

公共时间戳,则可结合极值点位置和采样的序列的趋势进行分组。并非所有公共时间戳对应的极值点都可作为分组的关键点,关键点的位置需要根据每一组序列提取的 shapelets 的判别力进行调整,shapelets 的判别力直观地体现为分类的精度,最后确定能使得 shapelets 判别力更高的关键点,进而最终确定分组。

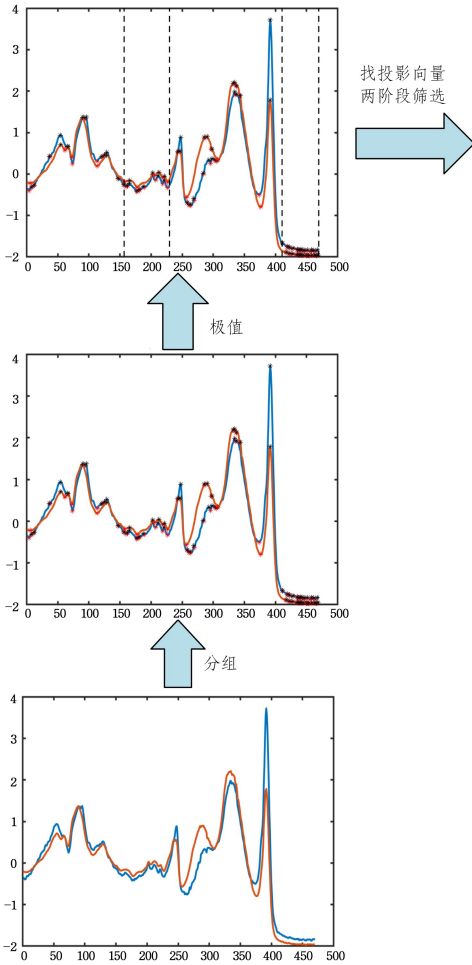


图3 数据集“Coffee”的分组结果

Fig. 3 Grouping results of dataset “Coffee”

3.1.3 系数连接矩阵

通过上述分组方法将全长为 l 的时间序列划分为每一组尺寸为 m_j 的 J 个组, $\sum_{j=1}^J m_j = l, \mathbf{v}^{(j)} \in \mathbb{R}^{m_j \times 1}$ 是每一组时间序

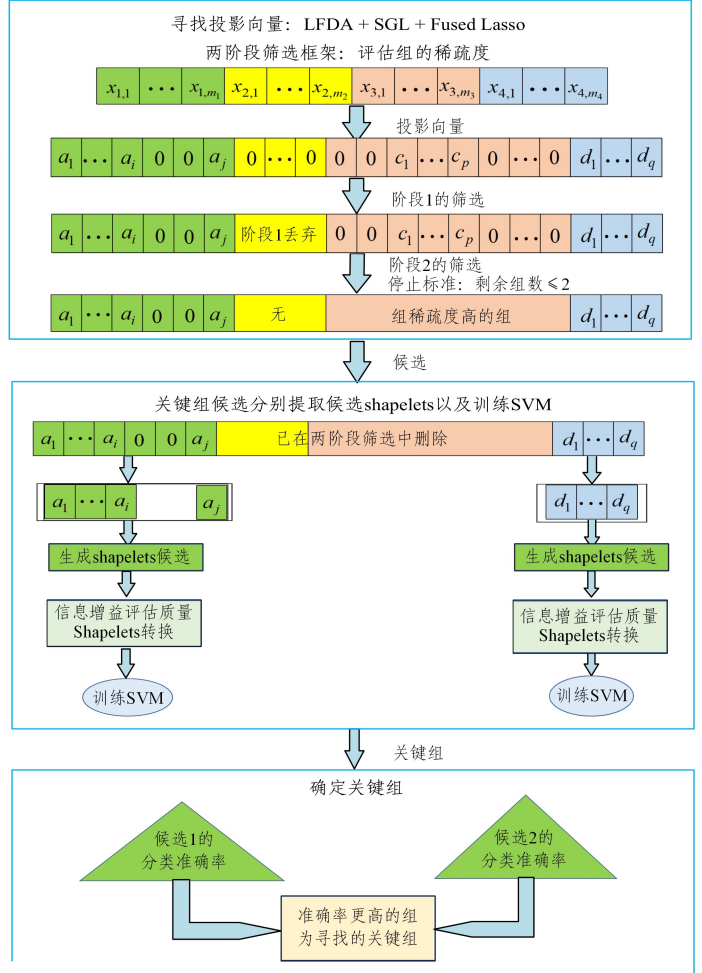


图2 SE2PF的框架图

Fig. 2 Framework of SE2PF

列对应的投影向量,构造每一组时间序列对应的系数矩阵 $\mathbf{v}_j \in \mathbb{R}^{m_j \times 1}, j = 1, 2, \dots, J$, 则:

列对应的投影向量,构造每一组时间序列对应的系数矩阵 $\mathbf{v}_j \in \mathbb{R}^{m_j \times 1}, j = 1, 2, \dots, J$, 则:

$$\mathbf{v} = (\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(J)})^T = (\mathbf{A}_1 \mathbf{v}, \mathbf{A}_2 \mathbf{v}, \dots, \mathbf{A}_J \mathbf{v})^T$$

对于 $j = 1, 2, \dots, J$, 可得各系数矩阵分别为:

$$\mathbf{A}_1 = (\mathbf{I}_{m_1}, \mathbf{O}_{m_1 \times (l-m_1)})_{m_1 \times l}$$

$$\mathbf{A}_2 = (\mathbf{O}_{m_2 \times m_1}, \mathbf{I}_{m_2}, \mathbf{O}_{l-m_1-m_2})_{m_2 \times l}$$

⋮

$$\mathbf{A}_i = (\mathbf{O}_{m_i \times (m_1+m_2+\dots+m_{i-1})}, \mathbf{I}_{m_i}, \mathbf{O}_{m_i \times (l-m_1-m_2-\dots-m_{i-1}-m_i)})_{m_i \times l}$$

⋮

$$\mathbf{A}_J = (\mathbf{O}_{m_J \times (l-m_J)}, \mathbf{I}_{m_J})_{m_J \times l}$$

(6)

其中, \mathbf{I} 为单位矩阵, \mathbf{O} 为零矩阵。

3.2 目标函数及求解

3.2.1 稀疏约束正则器

结合 3.1 节中的分组结果以及本文构建的系数连接矩阵,对稀疏组 Lasso 的每一组均赋予权重,加入融合 Lasso 的融合罚正则项,又 $\|\mathbf{v}\|_{2,1} = \sum_{j=1}^J \|\mathbf{v}^{(j)}\|_2$, 则本文的稀疏约束正则器为:

$$r(\mathbf{v}) = \alpha \sum_{j=1}^J \frac{\sqrt{m_j}}{l} \|\mathbf{A}_j \mathbf{v}\|_2 + \beta \|\mathbf{v}\|_1 + \eta \sum_{j=1}^J \|\mathbf{B}_j \mathbf{A}_j \mathbf{v}\|_1 \quad (7)$$

其中, 第一项保证组间的稀疏性; 第二项实现组内稀疏性, $\sqrt{m_j}/l$ 是对每一组赋予的权重, 在保证鲁棒性的同时确保不同组的相对重要程度; 第三项是在加权稀疏组 Lasso 的每一组内加入融合 Lasso 的融合罚正则项, 以保证每一组内变量的先后顺序, 同时使得解向量的相邻位置的值尽可能接近。

3.2.2 目标函数

本文将上述构造的稀疏约束正则器与局部线性判别分析相结合, 得到本文学习时间序列 shapelets 位置的稀疏模型, 目标函数如下:

$$f(\mathbf{v}) = \frac{\mathbf{v}^T \mathbf{S}_w \mathbf{v}}{\mathbf{v}^T \mathbf{S}_b \mathbf{v}} + r(\mathbf{v}) \quad (8)$$

可将其转化为式(9)所示的带约束的目标函数, 求得我们寻找的投影向量。

$$\min(\mathbf{v}^T \mathbf{S}_w \mathbf{v} + r(\mathbf{v})), \text{ s. t. } \mathbf{v}^T \mathbf{S}_b \mathbf{v} = 1 \quad (9)$$

其中, $\hat{\mathbf{S}}_b = \mathbf{S}_b + \tau \text{trace}(\mathbf{S}_b) \mathbf{I}/l$, 等价于对 \mathbf{S}_b 正则化^[35], 以避免用于训练的时间序列数量太少而导致其不满秩。

由于正则器的限制, 难以直接求解目标函数, 得到目标函数的最小值, 即求得所需投影向量, 因此本文采用交替方向乘法进行求解, 引入辅助向量 \mathbf{a}_j , \mathbf{f} 和 \mathbf{e}_j 帮助求解, 令 $\mathbf{a}_j = \mathbf{A}_j \mathbf{v}$, $\mathbf{f} = \mathbf{v}$, $\mathbf{e}_j = \mathbf{B}_j \mathbf{A}_j \mathbf{v}$, $j=1, 2, \dots, J$, 可得稀疏约束 $r(\mathbf{a}, \mathbf{f}, \mathbf{e})$, 则式(9)的增广拉格朗日式为:

$$\begin{aligned} L(\mathbf{v}, \mathbf{a}, \mathbf{e}, \mathbf{f}, \boldsymbol{\zeta}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = & \mathbf{v}^T \mathbf{S}_w \mathbf{v} + r(\mathbf{a}, \mathbf{f}, \mathbf{e}) + \lambda(1 - \mathbf{v}^T \hat{\mathbf{S}}_b \mathbf{v}) + \\ & \sum_{j=1}^J \zeta_j (\mathbf{A}_j \mathbf{v} - \mathbf{a}_j) + \boldsymbol{\mu} (\mathbf{v} - \mathbf{f}) + \\ & \sum_{j=1}^J \boldsymbol{\sigma}_j (\mathbf{B}_j \mathbf{A}_j \mathbf{v} - \mathbf{e}_j) + \frac{\rho_1}{2} \sum_{j=1}^J \|\mathbf{A}_j \mathbf{v} - \\ & \mathbf{a}_j\|_2^2 + \frac{\rho_2}{2} \|\mathbf{v} - \mathbf{f}\|_2^2 + \frac{\rho_3}{2} \sum_{j=1}^J \\ & \|\mathbf{B}_j \mathbf{A}_j \mathbf{v} - \mathbf{e}_j\|_2^2 \end{aligned} \quad (10)$$

3.2.3 目标函数的求解

(1) 固定 $\mathbf{a}_j, \mathbf{f}, \mathbf{e}_j, \boldsymbol{\zeta}_j, \boldsymbol{\mu}, \boldsymbol{\sigma}_j$, 更新 \mathbf{v} 。

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{v}} = & (\mathbf{S}_w + \mathbf{S}_w^T) \mathbf{v} - \lambda(\hat{\mathbf{S}}_b + \hat{\mathbf{S}}_b^T) \mathbf{v} + \sum_{j=1}^J \mathbf{A}_j^T \boldsymbol{\zeta}_j^T + \boldsymbol{\mu}^T + \sum_{j=1}^J \mathbf{A}_j^T \mathbf{B}_j^T \\ & \boldsymbol{\sigma}_j^T + \rho_1 \sum_{j=1}^J \mathbf{A}_j^T (\mathbf{A}_j \mathbf{v} - \mathbf{a}_j) + \rho_2 (\mathbf{v} - \mathbf{f}) + \\ & \rho_3 \sum_{j=1}^J \mathbf{A}_j^T \mathbf{B}_j^T (\mathbf{B}_j \mathbf{A}_j \mathbf{v} - \mathbf{e}_j) = 0 \end{aligned}$$

化简后令:

$$\begin{aligned} \mathbf{E} = & 2\mathbf{S}_w - 2\lambda\hat{\mathbf{S}}_b + \rho_1 \sum_{j=1}^J \mathbf{A}_j^T \mathbf{A}_j + \rho_2 \mathbf{I} + \rho_3 \sum_{j=1}^J \mathbf{A}_j^T \mathbf{B}_j^T \mathbf{B}_j \mathbf{A}_j \\ \mathbf{g} = & \rho_1 \sum_{j=1}^J \mathbf{A}_j^T \left(\mathbf{a}_j - \frac{\boldsymbol{\zeta}_j^T}{\rho_1} \right) + \rho_2 \left(\mathbf{f} - \frac{\boldsymbol{\mu}^T}{\rho_2} \right) + \rho_3 \sum_{j=1}^J \mathbf{A}_j^T \mathbf{B}_j^T \\ & \left(\mathbf{e}_j - \frac{\boldsymbol{\sigma}_j^T}{\rho_3} \right) \end{aligned}$$

再令:

$$\mathbf{R} = \mathbf{S}_w + \frac{1}{2} \rho_1 \sum_{j=1}^J \mathbf{A}_j^T \mathbf{A}_j + \frac{1}{2} \rho_2 \mathbf{I} + \frac{1}{2} \rho_3 \sum_{j=1}^J \mathbf{A}_j^T \mathbf{B}_j^T \mathbf{B}_j \mathbf{A}_j$$

结合 Cholesky 分解 $\hat{\mathbf{S}}_b = \mathbf{L}\mathbf{L}^T$, 则有:

$$\mathbf{E}^{-1} = \frac{1}{2} (\hat{\mathbf{S}}_b^{-1} \mathbf{R} \hat{\mathbf{S}}_b - \lambda \hat{\mathbf{S}}_b)^{-1}$$

$$= \frac{1}{2} \mathbf{L}^{-T} (\mathbf{L}^{-1} \mathbf{R} \mathbf{L}^{-T} - \lambda \mathbf{I})^{-1} \mathbf{L}^{-1}$$

若 $\mathbf{v}^t, \mathbf{a}^t, \mathbf{f}^t, \boldsymbol{\zeta}^t, \boldsymbol{\mu}^t$ 是第 t 次迭代的解, 则投影向量的第 $t+1$ 次迭代公式如下:

$$\begin{aligned} \mathbf{v}^{t+1} = & \frac{1}{2} \mathbf{L}^{-T} \left[\mathbf{L}^{-1} \left(\mathbf{S}_w + \frac{1}{2} \rho_1 \sum_{j=1}^J \mathbf{A}_j^T \mathbf{A}_j + \frac{1}{2} \rho_2 \mathbf{I} + \right. \right. \\ & \left. \left. \frac{1}{2} \rho_3 \sum_{j=1}^J \mathbf{A}_j^T \mathbf{B}_j^T \mathbf{B}_j \mathbf{A}_j \right) \mathbf{L}^{-T} - \lambda \mathbf{I} \right]^{-1} \mathbf{L}^{-1} \left[\rho_1 \sum_{j=1}^J \mathbf{A}_j^T \right. \\ & \left. \left(\mathbf{a}_j^t - \frac{\boldsymbol{\zeta}_j^t}{\rho_1} \right) + \rho_2 \left(\mathbf{f}^t - \frac{\boldsymbol{\mu}^t}{\rho_2} \right) + \rho_3 \sum_{j=1}^J \mathbf{A}_j^T \mathbf{B}_j^T \left(\mathbf{e}_j^t - \frac{\boldsymbol{\sigma}_j^t}{\rho_3} \right) \right] \end{aligned} \quad (11)$$

(2) 固定 \mathbf{v} 与 $\boldsymbol{\mu}$, 更新 \mathbf{f} 。

$$\frac{\partial L}{\partial \mathbf{f}} = \frac{\partial}{\partial \mathbf{f}} \left[\beta \|\mathbf{f}\|_1 + \boldsymbol{\mu}(\mathbf{v} - \mathbf{f}) + \frac{\rho_2}{2} \|\mathbf{v} - \mathbf{f}\|_2^2 \right] = 0$$

为了简化运算, 可将上式写作:

$$\frac{\partial L}{\partial \mathbf{f}} = \frac{\partial}{\partial \mathbf{f}} \left[\beta \|\mathbf{f}\|_1 + \frac{\rho_2}{2} \left\| \mathbf{v} - \mathbf{f} + \frac{\boldsymbol{\mu}}{\rho_2} \right\|_2^2 \right] = 0$$

则 \mathbf{f} 的第 $t+1$ 次迭代结果为:

$$\mathbf{f}^{t+1} = \begin{cases} \mathbf{v}^{t+1} + \frac{\boldsymbol{\mu}^t}{\rho_2} + \frac{\beta}{\rho_2}, & \mathbf{v}^{t+1} + \frac{\boldsymbol{\mu}^t}{\rho_2} < -\frac{\beta}{\rho_2} \\ \mathbf{v}^{t+1} + \frac{\boldsymbol{\mu}^t}{\rho_2} - \frac{\beta}{\rho_2}, & \mathbf{v}^{t+1} + \frac{\boldsymbol{\mu}^t}{\rho_2} > \frac{\beta}{\rho_2} \\ 0, & \left| \mathbf{v}^{t+1} + \frac{\boldsymbol{\mu}^t}{\rho_2} \right| < \frac{\beta}{\rho_2} \end{cases} \quad (12)$$

(3) 固定 \mathbf{v} 与 $\boldsymbol{\zeta}_j$, 更新 \mathbf{a}_j 。

对于 $j=1, 2, \dots, J$:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{a}_j} = & \frac{\partial}{\partial \mathbf{a}_j} \left[\alpha \frac{\sqrt{m_j}}{l} \|\mathbf{a}_j\|_2 + \frac{\rho_1}{2} \|\mathbf{A}_j \mathbf{v} - \mathbf{a}_j + \frac{\boldsymbol{\zeta}_j}{\rho_1}\|_2^2 \right] = 0 \\ & \left(\rho_1 + \frac{\alpha \sqrt{m_j}/l}{\|\mathbf{a}_j\|_2} \right) \mathbf{a}_j = \rho_1 \mathbf{A}_j \mathbf{v} + \boldsymbol{\zeta}_j \end{aligned}$$

由于 \mathbf{a}_j 不能显性表示, 令 $Z_j = \rho_1 + \frac{\alpha \sqrt{m_j}/l}{\|\mathbf{a}_j\|_2}$, 则有 $\mathbf{a}_j = \frac{1}{Z_j} (\rho_1 \mathbf{A}_j \mathbf{v} + \boldsymbol{\zeta}_j)$, 于是可得 Z_j 的表达式及 \mathbf{a}_j 的迭代结果:

$$\mathbf{a}_j^{t+1} = \frac{(\|\rho_1 \mathbf{A}_j \mathbf{v}^{t+1} + \boldsymbol{\zeta}_j\|_2 - \alpha \sqrt{m_j}/l) (\rho_1 \mathbf{A}_j \mathbf{v}^{t+1} + \boldsymbol{\zeta}_j)}{\rho_1 \|\rho_1 \mathbf{A}_j \mathbf{v}^{t+1} + \boldsymbol{\zeta}_j\|_2} \quad (13)$$

(4) 固定 \mathbf{v} 与 $\boldsymbol{\sigma}_j$, 更新 \mathbf{e}_j 。

$$\frac{\partial L}{\partial \mathbf{e}_j} = \frac{\partial}{\partial \mathbf{e}_j} \left[\eta \|\mathbf{e}_j\|_1 + \frac{\rho_3}{2} \left\| \mathbf{B}_j \mathbf{A}_j \mathbf{v} - \mathbf{e}_j + \frac{\boldsymbol{\sigma}_j}{\rho_3} \right\|_2^2 \right] = 0$$

则有:

$$\mathbf{e}_j^{t+1} = \begin{cases} \mathbf{B}_j \mathbf{A}_j \mathbf{v}^{t+1} + \frac{\boldsymbol{\sigma}_j^t}{\rho_3} + \frac{\eta}{\rho_3}, & \mathbf{B}_j \mathbf{A}_j \mathbf{v}^{t+1} + \frac{\boldsymbol{\sigma}_j^t}{\rho_3} < -\frac{\eta}{\rho_3} \\ \mathbf{B}_j \mathbf{A}_j \mathbf{v}^{t+1} + \frac{\boldsymbol{\sigma}_j^t}{\rho_3} - \frac{\eta}{\rho_3}, & \mathbf{B}_j \mathbf{A}_j \mathbf{v}^{t+1} + \frac{\boldsymbol{\sigma}_j^t}{\rho_3} > \frac{\eta}{\rho_3} \\ 0, & \left| \mathbf{B}_j \mathbf{A}_j \mathbf{v}^{t+1} + \frac{\boldsymbol{\sigma}_j^t}{\rho_3} \right| < \frac{\eta}{\rho_3} \end{cases} \quad (14)$$

(5) 分别固定 $\mathbf{v} \& \mathbf{a}_j, \mathbf{v} \& \mathbf{f}$ 和 $\mathbf{v} \& \mathbf{e}_j$, 更新拉格朗日乘子 $\boldsymbol{\zeta}_j$, $\boldsymbol{\mu}$ 和 $\boldsymbol{\sigma}_j$ 。

$$\boldsymbol{\zeta}_j^{t+1} = \boldsymbol{\zeta}_j^t + (\mathbf{A}_j \mathbf{v}^{t+1} - \mathbf{a}_j^{t+1}) \quad (15)$$

$$\boldsymbol{\mu}^{t+1} = \boldsymbol{\mu}^t + (\mathbf{v}^{t+1} - \mathbf{f}^{t+1}) \quad (16)$$

$$\boldsymbol{\sigma}_j^{t+1} = \boldsymbol{\sigma}_j^t + (\mathbf{B}_j \mathbf{A}_j \mathbf{v}^{t+1} - \mathbf{e}_j^{t+1}) \quad (17)$$

通过迭代更新优化目标函数, 直到满足收敛条件或达到最大迭代次数求得投影向量, 具体优化过程如算法 1 所示。

算法 1 SE2PF 投影向量的优化算法

输入:根据分组结果输入系数连接矩阵 A_j , 稀疏矩阵 $B_j, j=1, 2, \dots, J$, 稀疏约束参数 α, β, η , 惩罚参数 ρ_1, ρ_2, ρ_3 , 最大迭代次数 T 及停止阈值 γ

输出:输出每一组对应的投影向量 $v^{(j)}$

初始化:投影向量 v , 辅助向量 f, a_j, e_j , 拉格朗日乘子 $\zeta_j, \mu, \sigma_j, \lambda$;

1. 计算局部类内离散度矩阵 S_w 和局部类间离散度矩阵 S_b ;
2. for $t=1, 2, \dots, T$ do
3. Repeat
4. 根据式(11)迭代更新投影向量 v ;
5. 根据式(12)更新辅助向量 f^{t+1} ;
6. 根据式(13)更新 a_j^{t+1} ;
7. 根据式(14)更新 e_j^{t+1} ;

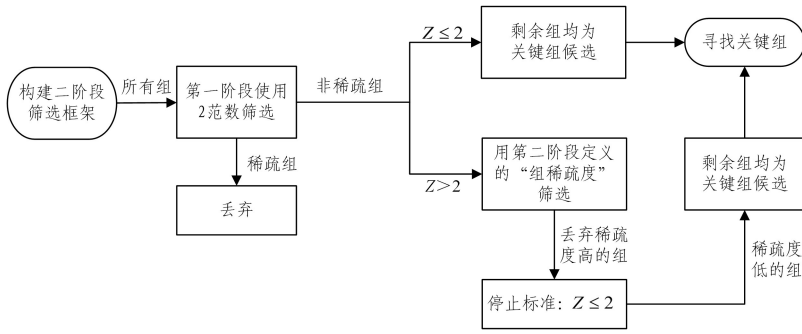


图 4 两阶段筛选框架

Fig. 4 Two-phase filtering framework

第一阶段 使用 l_2 范数筛选稀疏组和非稀疏组, 根据每个数据集的具体情况, 将每个组的投影向量 $v^{(j)}$ 的 l_2 范数值与 $s = \{0.1, 0.01, 0.001\}$ 中的值进行比较, 以此区分稀疏组和非稀疏组。

$$\begin{cases} \|v^{(j)}\|_2 \leq s, & j \text{ 为稀疏组} \\ \|v^{(j)}\|_2 > s, & j \text{ 为非稀疏组} \end{cases} \quad (18)$$

其中, $j=1, 2, \dots, J$ 。经筛选框架的第一阶段过滤后, 丢弃稀疏组, 非稀疏组进入第二阶段。

第二阶段 定义“组稀疏度”进行二次过滤。一般地, 经第一阶段筛选后, 此时剩余组数 $Z > 2$ 较多, 如果每一组都分别用于提取 shapelets, 那么寻找关键组的时间成本就会过高, 因此定义式(19)的“组稀疏度”进行二次过滤, 提前丢弃高稀疏度的组, 低稀疏度的组可视为关键组候选, 作为提取 shapelets 的候选组。

$$GSD(j) = \frac{L(n_j)}{L(m_j)}, \quad j=1, 2, \dots, J \quad (19)$$

其中, $L(\cdot)$ 表示长度, m_j 是每组的长度, n_j 表示第 j 组中“零”的个数, 根据每组对应的投影向量的实际情况, 从集合 $s = \{0.1, 0.01, 0.001\}$ 中选择合适的值与投影向量的绝对值进行比较, 以确定“零”的个数, 对同一个时间序列数据集所有组的评估标准应保持一致。非零项实际对应 shapelets 候选项的位置, 因此丢弃零项比例高的组, 寻找非零项比例高的组, 即保留稀疏度低的组, 提前丢弃稀疏度高的组, 具有一定的理论基础。筛选框架的第二阶段设置的停止标准为剩余组 Z 仅为两组, 此时筛选框架即可停止。两阶段筛选框架的具体过滤过程如算法 2 所示。

算法 2 SE2PF 关键组的快速定位过滤算法

输入:所有组的投影向量 $v^{(j)}, j=1, 2, \dots, J$

输出:分类准确率最高的组即关键组 h , 对应的分类准确率与运行时间

8. 根据式(15)更新拉格朗日乘子 ζ_j^{t+1} ;
9. 根据式(16)更新 μ^{t+1} ;
10. 根据式(17)更新 σ_j^{t+1} ;
11. Until $\|old_v - v\| < \gamma$, 迭代停止;
12. end for

3.3 两阶段筛选框架

前述工作中已获得每组的投影向量 $v^{(j)}$, 本文提出两阶段筛选框架的目的是更快地定位关键组, 仅使用关键组提取 shapelets 以及训练分类器 SVM, 进而对测试集进行分类。特别地, 本文选择 SVM 作为分类器是因为其成熟度以及近年来一些与 shapelets 相关的文章^[11, 22, 25]的分类均在 SVM 分类器的帮助下顺利完成。筛选框架如图 4 所示。

1. 第一阶段:根据式(18)使用 l_2 范数区分稀疏组与非稀疏组, 提前将稀疏组丢弃;
2. 非稀疏组进入筛选框架的第二阶段;
3. 第二阶段:根据式(19)评估进入第二阶段的组的稀疏度 $GSD(j)$;
4. SORT $GSD(j)$;
5. 重复丢弃稀疏度最高的组;
6. Until $Z \leq 2$, 筛选停止;
7. 剩余至多两组为关键组候选, 分别使用关键组候选提取 shapelets, 训练 SVM, 对测试集进行分类。

3.4 从 shapelets 指示器获取 shapelets

经两阶段筛选框架的过滤后, 保留至多两个剩余组, 均为关键组候选。每一组的非零位置即 shapelets 指示器所在位置。由 shapelets 指示器获得 shapelets 候选, 经信息增益评估获得 shapelets, 最后利用 shapelets 将原始时间序列转换为特征向量^[11], 目的是寻找与 shapelet 最相似的子序列。第 p 条时间序列 X_p 与第 q 个 shapelets F_q 的最小距离定义如下:

$$distance_{p,q} = \min_{h=1, 2, \dots, H} \frac{1}{S} \sum_{s=1}^S (X_{p, h+s-1} - F_{q,s})^2 \quad (20)$$

其中, H 表示长度为 $l-S+1$ 的 X_p 的段数, S 是 F_q 的长度。本文仅用一组提取 shapelets, 两组关键组候选的 shapelets 提取以及对时间序列的分类是独立的, 最终只确定一组关键组。

4 实验验证与分析

本文的实验环境为 MATLAB 2016b, 64 位操作系统的 Windows 10 中文版, Intel i7 CPU 以及 16GB 的内存。

4.1 数据集

为了便于比较, 本文以 FLAG 与 SEDS 等方法为依据, 从 UCR 和 UEA 时间序列数据库中选取完全相同的 28 个一元时间序列数据集作为研究对象, 其基本信息如表 2 所列。

表2 本文实验使用的数据集

Table 2 Datasets used in our experiments

数据集	训练数	测试数	长度	类别数	数据集	训练数	测试数	长度	类别数
Adiac	390	391	176	37	Lighting7	70	73	319	7
Beef	30	30	470	5	Medical	381	760	99	10
BeetleFly	20	20	512	2	MoteStrain	20	1252	84	2
BirdChicken	20	20	512	2	MP_Little	400	645	250	3
Chlorine	467	3840	166	3	MP_Middle	400	645	250	3
Coffee	28	28	286	2	Otoliths	64	64	512	2
Diatom	16	306	345	4	PP_Little	400	645	250	3
DP_Little	400	645	250	3	PP_Middle	400	645	250	3
DP_Middle	400	645	250	3	PP_Thumb	400	645	250	3
DP_Thumb	400	645	250	3	Sony	20	601	70	2
ECGFive	23	861	136	2	Symbols	25	995	398	6
FaceFour	24	88	350	4	Synthetic	300	300	60	6
Gun_Point	50	150	150	2	Trace	100	100	275	4
ItalyPower	67	1029	24	2	TwoLead	23	1139	82	2

表2 数据类型丰富,覆盖多个领域,包含8个图像数据集、6个传感器数据集、2个光谱数据集、2个心电图数据集、1个运动数据集、1个模拟数据集以及8个骨骼数据集。

4.2 实验对照组

为了便于比较,从FLAG与SEDS论文中选取以下8个基于shapelets的时间序列分类方法作为对照组。

(1)IG^[8]:一种使用信息增益度量shapelets的质量的基于shapelets的分类方法。

(2)FS^[11]:一种使用F统计量度量shapelets的质量的基于shapelets的分类方法。

(3)KW^[9]:使用Kruskal-Wallis检验shapelets的质量的基于shapelets的分类方法。

(4)IGSVM^[11]:一种基于支持向量机和信息增益的

shapelet变换算法。

(5)LTS^[21]:一种使用随机梯度下降法求解逻辑损失函数从而学习近似最优的shapelets的分类方法。

(6)UFS^[18]:一种引入随机shapelets的分类方法。

(7)FLAG^[22]:一种使用广义特征向量法和融合Lasso正则器来学习shapelets的位置的分类方法。

(8)SEDS^[23]:一种使用局部线性判别分析和融合Lasso正则器来学习shapelets的位置的分类方法。

4.3 分组与参数

本文采用极值关键点采样的训练时间序列进行分组,3.1节中详细介绍了这种分组方法。

表3列出了实验中28个数据集的具体分组信息以及参数信息。

表3 每个数据集的分组信息和参数信息

Table 3 Group information and parameter information of each dataset

数据集	分组信息		参数信息			
	组数	各组长度	α	β	η	k
Adiac	3	41/96/39	0.700	0.25000	0.0900	5
Beef	6	74/75/53/52/15/201	0.200	0.07500	0.0015	2
BeetleFly	5	64/33/61/78/276	0.200	0.00350	0.0200	3
BirdChicken	5	31/54/119/248/60	0.450	0.00800	0.1000	2
Chlorine	2	151/15	0.100	0.08000	0.4500	4
Coffee	4	104/32/45/105	0.150	0.00705	0.1500	3
Diatom	2	332/13	0.350	0.90000	0.1500	3
DP_Little	2	22/228	0.080	0.01500	0.0200	9
DP_Middle	2	241/9	0.040	0.80000	0.1500	9
DP_Thumb	2	211/39	0.800	0.70000	0.1000	9
ECGFive	3	13/113/10	0.650	0.00350	0.0150	3
FaceFour	4	42/32/137/139	0.450	0.95000	0.0550	3
Gun_Point	4	64/24/34/28	0.400	0.15000	0.0050	3
ItalyPower	2	6/18	0.900	0.00200	0.1500	7
Lighting7	5	61/52/87/63/56	0.250	0.85000	0.1000	5
Medical	3	21/18/60	0.700	0.80000	0.1000	10
MoteStrain	2	81/3	0.030	0.00250	0.0025	6
MP_Little	3	14/210/26	0.040	0.40000	0.3000	9
MP_Middle	3	10/220/20	0.090	0.50000	0.0020	13
Otoliths	3	118/321/73	0.750	0.01500	0.0100	3
PP_Little	3	157/38/55	0.500	0.60000	0.0400	9
PP_Middle	2	220/30	0.800	0.30000	0.0200	9
PP_Thumb	2	239/11	0.009	0.00500	0.3000	11
Sony	3	8/58/4	0.850	0.00300	0.0025	5
Symbols	3	16/351/31	0.100	0.50000	0.1000	2
Synthetic	4	10/7/24/19	0.015	0.00150	0.0150	3
Trace	3	97/128/50	0.350	0.00200	0.0010	5
TwoLead	2	65/17	0.060	0.35000	0.5500	3

每个数据集对应的具体的分组数以及分组后每个组的具体长度如表 3 所列,加粗数字表示实验筛选出的最终用于 shapelets 提取的分类关键组对应的长度。例如,“Beef”划分为 6 组,具体的分组长度分别为 74-75-53-52-15-201,且实验中最终确定的关键组为第 6 组,序列长度为 201。

此外,表 3 还列出了与实验相关的 3 个超参数 α, β, η 和用于局部缩放的局部参数 k 。其中, k 通过交叉验证获取, 3 个稀疏约束参数 α, β, η 是对训练数据集进行交叉验证并结合网格搜索得到的最优参数组合。具体地,每个稀疏约束参数有 54 个参数值候选,分别是在区间 $(0.1, 0.95)$, $(0.01, 0.095)$, $(0.001, 0.0095)$ 上以 0.5, 0.05, 0.005 为间距取值,自由组合,并保留每一次的分类结果,最终获取最高分类准确率所对应的参数组合即为该数据集寻找的最优参数组合。

4.4 两阶段筛选框架

表 4 中每个数据集括号里的数字表示对应的分组的数量,“稀疏组”列表示筛选框架的第一阶段筛选出的稀疏组,

这些被筛选出的稀疏组不参与框架第二阶段的筛选。符号“—”表示筛选框架的第一阶段未筛选出稀疏组;符号“/”表示该组为稀疏组,已在筛选框架的第一阶段被丢弃;符号“#”表示数据集在经第一阶筛选后只剩下一组,因此不需要使用筛选框架的第二阶段,可以直接将其看作该数据集的关键组,用于 shapelets 的提取以及 SVM 模型的训练。另外,二阶段加粗的数据为筛选框架的第二阶段丢弃的组对应的 GSD 值。例如,数据集“BeetleFly”共分为 5 组,筛选框架的第一阶段判断第 2 组为稀疏组,丢弃第 2 组,第 1,3,4 和 5 组进入筛选框架的第二阶段,组 1 和组 4 的稀疏度高,丢弃,不参与时间序列关键组的寻找,最后剩余组 3 和组 5 为关键组的候选,分别使用这两组提取 shapelets,训练 SVM,使用训练好的 SVM 对测试集进行分类,最终比较分类准确率后确定关键组为组 3。数据集“Coffee”共分为 4 组,第一阶段判断其组 1 和组 2 为稀疏组,剩余的组 3 和组 4 被直接视为关键组的候选,不再使用筛选框架的第二阶段评估这两组的稀疏度,最终确定关键组为组 4。

表 4 每个数据集的两阶段筛选信息

Table 4 Two phase filtering information of each dataset

数据集	稀疏组	二阶						关键组
		1	2	3	4	5	6	
Adiac(3)	—	0.220	0.135	0.359				1
Beef(6)	—	0.378	0.400	0.679	0.346	0.067	0.309	6
BeetleFly(5)	2	0.781	/	0.689	0.782	0.457		3
BirdChicken(5)	1,2	/	/	0.656	0.432	0.300		4
Chlorine(2)	—	*	*					1
Coffee(4)	1,2	/	/	*	*			4
Diatom(2)	2	#	/					1
DP_Little(2)	—	*	*					2
DP_Middle(2)	—	*	*					1
DP_Thumb(2)	—	*	*					1
ECGFive(3)	1,3	/	#	/				2
FaceFour(4)	—	0.119	0.219	0.672	0.187			4
Gun_Point(4)	3,4	*	*	/	/			1
ItalyPower(2)	1	/	#					2
Lighting7(5)	—	0.246	0.404	0.230	0.079	0.286		3
Medical(3)	—	0.048	0.056	0.000				1
MoteStrain(2)	2	#	/					1
MP_Little(3)	—	0.485	0.454	0.458				2
MP_Middle(3)	—	0.242	0.212	0.256				2
Otoliths(3)	3	0.822	0.561	/				1
PP_Little(3)	—	0.213	0.445	0.224				1
PP_Middle(2)	—	*	*					1
PP_Thumb(2)	—	*	*					1
Sony(3)	3	0.000	0.224	/				2
Symbols(3)	—	0.000	0.530	0.645				2
Synthetic(4)	—	0.000	0.000	0.000	0.000			3
Trace(3)	—	0.041	0.000	0.040				2
TwoLead(2)	2	#	/					1

4.5 实验结果对比

4.5.1 准确率对比

表 5 列出了对照组中 8 种基于 shapelets 的分类方法与本文提出的方法 SE2PF 的分类准确率对比,最后一列为 SE2PF 使用筛选出的关键组以及关键组对应的最佳参数下的分类准确性。其中“—”表示相应的分类方法无法在合理的时间内(即超过 24h)获得这些数据集的分类结果。每个数据集的最高准确率加粗表示,括号中的数字表示每个数据集

分类精度的排名,数字 1 表示分类准确率最高。结果表明,SE2PF 在实验中所使用的 28 个数据集上的 15 个数据集上的分类准确率排名第 1。28 个数据集的平均秩为 1.8,表明本文方法 SE2PF 在分类精度上优于所有基于 shapelets 的方法,也说明 SE2PF 提取出的 shapelets 与其他分类方法获得的 shapelets 相比具有更好的判别力。从平均秩的数值来看,针对本实验选取的 28 个公共数据集,分类精度从高到低排名依次为 SE2PF,SEDS,LTS,FLAG,IGSVM,FSH,FS,IG,KW。

表 5 准确率对比

Table 5 Accuracy comparison

(单位: %)

数据集	IG	FS	KW	FSH	IGSVM	LTS	FLAG	SEDS	SE2PF
Adiac	29.9(6)	15.6(9)	26.2(7)	57.5(4)	23.5(8)	51.9(5)	75.2(2)	77.2(1)	66.5(3)
Beef	50.0(7)	56.7(6)	33.3(8)	50.0(7)	90.0(2)	76.7(5)	83.3(4)	86.7(3)	93.3(1)
BeetleFly	77.5(4)	90.0(3)	70.0(5)	65.0(6)	97.5(1)	95.0(2)	90.0(3)	90.0(3)	95.0(2)
BirdChicken	85.0(5)	90.0(3)	87.5(4)	90.0(3)	95.0(2)	100.0(1)	85.0(5)	100.0(1)	95.0(2)
Chlorine	58.8(5)	53.5(7)	52.0(8)	58.8(5)	57.1(6)	73.0(4)	76.0(3)	77.9(2)	78.0(1)
Coffee	96.4(2)	100.0(1)	85.7(4)	92.9(3)	100.0(1)	100.0(1)	100.0(1)	100.0(1)	100.0(1)
Diatom	76.5(6)	76.5(6)	62.1(7)	87.3(5)	93.1(4)	94.2(3)	96.4(2)	98.7(1)	98.7(1)
DP_Little	—	—	—	60.6(6)	66.6(3)	73.4(1)	68.3(2)	64.5(5)	65.9(4)
DP_Middle	—	—	—	58.8(6)	69.5(5)	74.1(2)	71.3(4)	73.6(3)	75.7(1)
DP_Thumb	—	—	—	63.4(6)	69.6(5)	75.2(1)	70.5(4)	73.3(3)	74.1(2)
ECGFive	77.5(6)	99.0(3)	87.2(5)	99.8(2)	99.0(3)	100.0(1)	92.0(4)	100.0(1)	100.0(1)
FaceFour	84.0(7)	75.0(8)	44.3(9)	92.0(5)	97.7(2)	94.3(4)	90.9(6)	96.6(3)	98.9(1)
Gun_Point	89.3(7)	95.0(3)	94.0(6)	94.0(6)	100.0(1)	99.6(2)	96.7(4)	100.0(1)	98.7(3)
ItalyPower	89.2(8)	93.1(6)	91.0(7)	91.0(7)	93.7(5)	95.8(3)	94.6(4)	96.1(2)	96.6(1)
Lighting7	49.3(7)	41.1(9)	48.0(8)	65.2(5)	63.0(6)	79.0(2)	76.7(3)	79.5(1)	71.2(4)
Medical	48.8(8)	50.8(7)	47.1(9)	64.7(5)	52.2(6)	71.3(2)	71.4(1)	71.1(3)	69.7(4)
MoteStrain	82.5(8)	84.0(6)	84.0(6)	83.8(7)	88.7(5)	90.0(2)	88.8(4)	89.5(3)	90.3(1)
MP_Little	—	—	—	56.9(6)	70.7(4)	74.3(2)	69.3(5)	73.2(3)	75.7(1)
MP_Middle	—	—	—	60.3(6)	76.9(2)	77.5(1)	75.0(4)	74.4(5)	75.5(3)
Otoliths	67.2(3)	57.8(7)	60.9(5)	60.9(5)	64.1(4)	59.4(6)	64.1(4)	75.0(2)	78.1(1)
PP_Little	—	—	—	57.6(6)	72.1(1)	71.0(3)	67.1(5)	69.3(4)	71.2(2)
PP_Middle	—	—	—	61.6(6)	75.9(2)	74.9(4)	73.8(5)	75.2(3)	76.9(1)
PP_Thumb	—	—	—	55.8(6)	75.5(1)	70.5(3)	67.4(5)	69.5(4)	70.7(2)
Sony	85.7(7)	95.3(3)	72.7(8)	68.6(9)	92.7(5)	91.0(6)	92.9(4)	96.2(2)	97.2(1)
Symbols	78.4(8)	90.1(5)	55.7(9)	92.4(2)	84.6(7)	94.5(1)	87.5(6)	91.3(4)	92.0(3)
Synthetic	94.3(7)	95.7(5)	90.0(8)	94.7(6)	87.3(9)	97.3(4)	99.7(1)	97.7(3)	99.0(2)
Trace	98.0(3)	100.0(1)	94.0(4)	100.0(1)	98.0(3)	100.0(1)	99.0(2)	100.0(1)	100.0(1)
TwoLead	85.1(6)	97.0(4)	76.4(7)	92.5(5)	100.0(1)	100.0(1)	99.0(3)	99.8(2)	100.0(1)
平均秩	6.0	5.2	6.7	5.2	3.7	2.6	3.6	2.5	1.8

4.5.2 时间复杂度对比

结合前人的研究可知, FLAG 和 SEDS 是目前公认的基于学习方式提取 shapelets 分类方法中运行速度较快的方法, 这两种方法比对照组中其他 6 种方法快很多。因此, 本文提出的 SE2PF 仅与这两种方法进行了比较, 以此验证本文方法在运行速率上的有效性。先对比 3 种方法的时间复杂度。假设时间序列的长度是 l , J 是每个数据集对应的组数, T 是交替方向乘法迭代的次数, n 是时间序列所有类别的总数。对于 J 类的数据集, FLAG 的算法复杂度是 $O(C(l^3 + Tl^2))$, SEDS 的算法复杂度是 $O(C(n^3 + Tl^2))$ 。本文方法 SE2PF 计算离散度矩阵的时间复杂度是 $O(nl^2)$, 交替方向乘法迭代更新的时间复杂度是 $O(C(n^3 + T(l + Jl + Jd)^2))$, 特别地, 当 $C < 5$ 时, $nl^2 \ll C(n^3 + T(l + Jl + Jd)^2)$, 因此 SE2PF 的时间复杂度是 $O(C(n^3 + T(l + Jl + Jd)^2))$ 。与 FLAG 和 SEDS 相比, 本文的 SE2PF 因存在分组而时间复杂度相对较高。

以上 3 种分类方法是在相同的计算机配置和环境下进行的, 具体的运行时间结果如表 6 所列。从表 6 的数据及平均秩的数值来看, FLAG 在几乎所有的数据集上分类最快, SE2PF 与 SEDS 的分类运行时间大致相同。另外, 结合时间复杂度和表 6 的运行时间以及表 5 的分类精度可得出以下结论: 1) 当实际应用中需要高精度的分类结果时, 可优先考虑 SE2PF; 2) SE2PF 在二分类问题中运行较快, 因此当实际的分类问题是二分类时, 可以优先使用 SE2PF; 3) SE2PF 在传感器数据集和心电图数据集上运行更快, 与 FLAG 和 SEDS 相比差异不大, 因此如果待分类数据集为这两种类型, 可优先考虑 SE2PF, 原因是其可以确保高分类精度又不会显著降低

分类速度。由此可知, 本文提出的 SE2PF 在不显著降低分类速度的前提下能提高分类精度, 证明了本文方法的有效性。

表 6 运行时间对比

Table 6 Running time comparison

(单位: s)

数据集	FLAG	SEDS	SE2PF
Adiac	1.50(1)	6.08(2)	10.61(3)
Beef	1.11(1)	4.39(2)	7.32(3)
BeetleFly	0.43(1)	1.27(2)	1.52(3)
BirdChicken	0.51(1)	1.09(3)	0.93(2)
Chlorine	0.17(1)	24.36(3)	16.02(2)
Coffee	0.11(1)	0.18(2)	0.36(3)
Diatom	0.34(1)	1.04(2)	2.39(3)
DP_Little	0.24(1)	5.62(2)	7.61(3)
DP_Middle	0.19(1)	6.67(2)	27.15(3)
DP_Thumb	0.27(1)	9.45(2)	31.77(3)
ECGFive	0.04(1)	0.32(3)	0.29(2)
FaceFour	0.23(1)	0.90(2)	2.59(3)
Gun_Point	0.04(1)	0.19(3)	0.15(2)
ItalyPower	0.01(1)	0.14(3)	0.11(2)
Lighting7	0.26(1)	1.73(2)	5.31(3)
Medical	0.20(1)	9.55(3)	2.66(2)
MoteStrain	0.02(1)	0.25(3)	0.21(2)
MP_Little	0.22(1)	9.12(2)	22.48(3)
MP_Middle	0.20(1)	5.92(2)	24.77(3)
Otoliths	1.18(2)	0.48(1)	1.40(3)
PP_Little	0.27(1)	9.03(2)	17.53(3)
PP_Middle	0.21(1)	6.23(3)	5.54(2)
PP_Thumb	0.25(1)	9.12(2)	20.31(3)
Sony	0.02(1)	0.13(3)	0.10(2)
Symbols	0.71(1)	4.29(2)	8.99(3)
Synthetic	0.07(1)	0.88(3)	0.85(2)
Trace	0.28(1)	1.25(2)	2.49(3)
TwoLead	0.05(1)	0.18(3)	0.15(2)
平均秩	1.04	2.36	2.61

4.6 shapelets 的规模

本文最终仅使用两阶段筛选框架辅助定位的一组关键字来提取 shapelets。现用实例证明所构建的 shapelets 学习模型能在一定程度上减小 shapelets 的规模。由表 5 结果可知,SEDS 的分类准确率排名仅次于 SE2PF,将这两种方法的 shapelets 数量进行直观的对比。对于数据集“Face-Four”,SE2PF 将 shapelets 的规模从 344 缩减到了 280;对于数据集“Medical”,SE2PF 将 shapelets 的规模从 661 缩小到了 416。由以上两个实例可知,本文提出的新的 shapelets 学习模型能够一定程度地减小 shapelets 的规模,促进分类进程。

4.7 参数灵敏度分析

参数的灵敏度可以用于分析与验证模型的稳定性。每个数据集有 3 个超参数,此处将 3 个参数中的一个固定

为最优参数组合中的最优值,其他两个参数结合最优取值,确定参数的变化范围,绘制参数灵敏度图。本文选取“Medical”和“Symbols”两个数据集,结合表 3 可知,“Medical”数据集的最优参数组合为 $\alpha=0.7, \beta=0.8, \eta=0.1$,“Symbols”数据集的最优参数组合是 $\alpha=0.1, \beta=0.5, \eta=0.1$ 。由图 5 与图 6 可知,参数的确定对不同数据集的分类性能有着不同的影响。参数的不同组合使得分类准确率结果随之发生变化。但具体来说,由图 5 可知,对于数据集“Medical”,准确率比较稳定,几乎都在 67.4%~69.7%之间稳定变化,变化幅度较小,证明该方法稳定性好;由图 6 可知,对于数据集“Symbols”,准确率同样比较稳定,几乎都在 84.4%~92.0%之间变化,变化幅度也较小,稳定性好。由此可证明本文提出的 SE2PF 方法稳定性好,模型相对比较稳定。

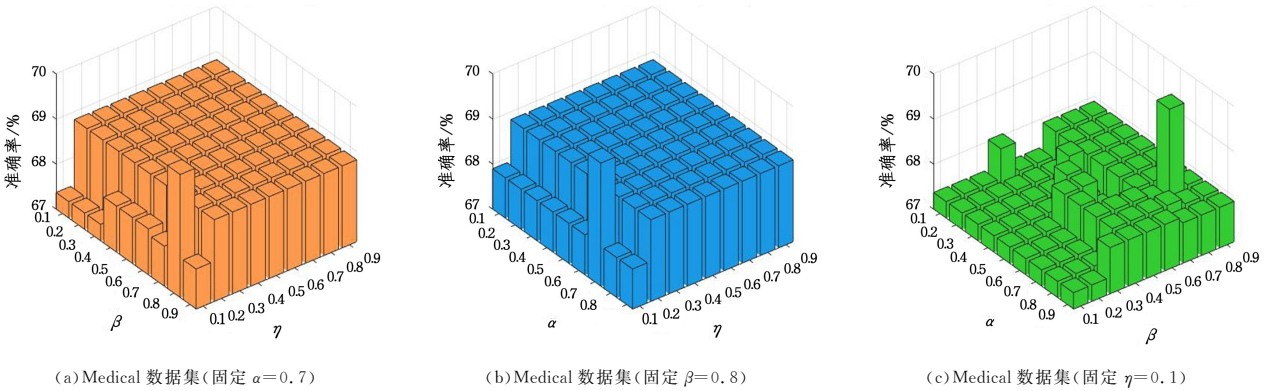


图 5 “Medical”数据集的 3 个参数的灵敏度分析图

Fig. 5 Sensitivity analysis graph of three parameters on “Medical” dataset

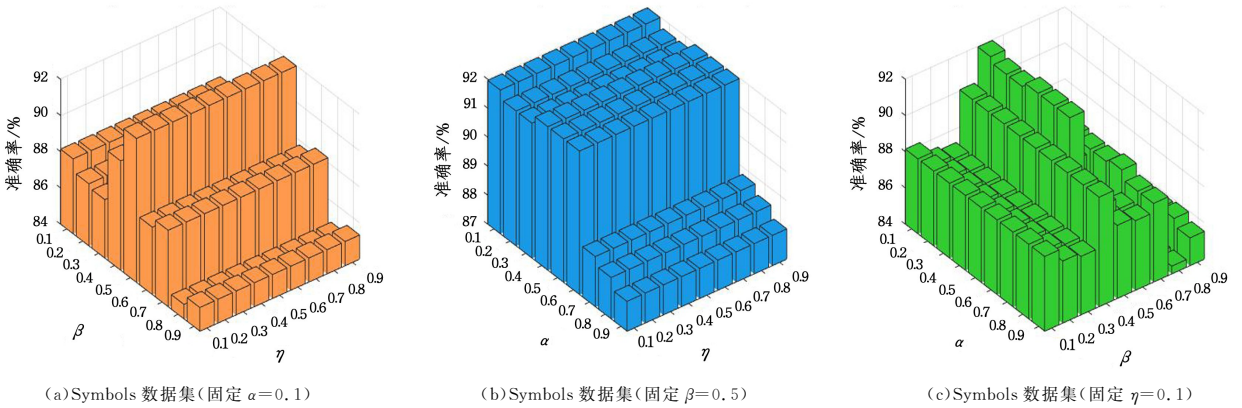


图 6 “Symbols”数据集的 3 个参数的灵敏度分析图

Fig. 6 Sensitivity analysis graph of three parameters on “Symbols” dataset

另外,为了更加全面地探究 3 个超参数值对所构建的 SE2PF 模型分类准确性的影响,本文结合表 3 中通过模型训练确定的每个数据集的最优超参数组合,进行了较为全面的参数灵敏度测试实验,结果如表 7 所列。

表 7 中每一列的含义是固定另外两个超参数为表 3 中对应的最优参数值,探究该参数的 54 个候选值的分类准确率及其变化范围,其结果用均值±标准差表示,以此分别研究每一个超参数在 SE2PF 模型中的灵敏度。

例如,对于数据集“Adiac”,由表 3 可知该数据集的最优参数组合是 $\alpha=0.7, \beta=0.25$ 以及 $\eta=0.09$ 。现在固定 $\beta=0.25$ 与 $\eta=0.09$,获取 α 变化时 SE2PF 模型的所有分类准确率,计算这些准确率的均值以及标准差,并将其表示为均值±标准差的形式。由表 7 也可看出,从整体来看,当固定另外两个超参数时,剩余的参数在 SE2PF 方法中分类精度变化幅度小,灵敏度不高,这也证明了 SE2PF 模型的鲁棒性。

表 7 参数灵敏度分析结果

Table 7 Parameter sensitivity analysis results

数据集	α	β	η	数据集	α	β	η
Adiac	65.45±0.50	62.71±1.51	63.85±1.61	Lighting7	69.84±3.23	65.30±4.23	65.95±2.88
Beef	91.36±1.65	87.50±5.95	92.73±2.16	Medical	67.93±0.25	67.69±0.54	68.17±0.58
BeetleFly	87.14±5.25	85.00±8.66	82.00±7.48	MoteStrain	88.67±1.48	84.86±4.50	86.09±1.69
BirdChicken	90.83±4.49	90.00±4.08	86.88±3.00	MP_Little	73.29±2.84	71.37±3.81	69.71±4.64
Chlorine	77.75±0.43	70.70±5.07	73.85±2.91	MP_Middle	69.95±4.38	71.84±3.15	72.88±3.02
Coffee	92.46±4.74	95.24±4.12	93.07±1.22	Otoliths	74.40±4.48	71.56±3.48	71.13±4.19
Diatom	98.44±0.14	93.34±2.72	94.35±2.00	PP_Little	69.54±2.52	67.63±4.15	70.26±1.53
DP_Little	62.24±3.57	60.55±4.21	61.32±3.26	PP_Middle	72.82±3.94	69.88±4.81	74.04±1.45
DP_Middle	72.30±2.95	73.54±2.11	70.15±5.54	PP_Thumb	68.21±3.25	66.82±5.07	68.61±2.38
DP_Thumb	72.11±2.18	69.85±3.15	71.18±3.08	Sony	92.85±2.43	89.18±4.30	90.45±4.08
ECGFive	94.84±8.29	91.75±5.39	92.55±4.46	Symbols	91.54±1.11	88.09±1.29	91.96±0.43
FaceFour	97.62±0.80	94.55±3.16	94.23±2.26	Synthetic	98.38±0.34	98.32±0.25	98.48±0.44
Gun_Point	97.63±0.88	94.31±1.80	94.49±2.79	Trace	98.07±0.67	96.26±3.21	97.98±0.42
ItalyPower	94.19±1.84	90.76±4.75	93.47±0.84	TwoLead	97.73±2.24	93.17±2.79	96.81±1.63

结束语 针对现有基于学习的模型提取的 shapelets 判别力不足, shapelets 候选规模仍然较大以及模型可解释性欠缺等问题, 本文提出了基于优化和两阶段筛选的时间序列 shapelets 提取算法 SE2PF。结合取样的时间序列的极值点位置及序列的趋势对其进行分组, 根据分组结果对稀疏组 Lasso 正则器的每项赋予权重, 构成加权 SGL 正则项, 将融合 Lasso 的融合罚具有的使解分段常数化的特性融入加权 SGL 正则项中, 组成新的正则器, 再与 LFDA 结合构建新的 shapelets 学习模型。此外借助两阶段筛选框架加速关键组的定位, 快速获取小规模且具有判别力的 shapelets。实验结果表明, 本文提出的两阶段筛选框架对关键组的快速定位是有效的, 同时构建的模型在有效减少 shapelets 数量的同时还能保证高判别力, 此外, 其分类速率高于大部分基于 shapelets 的分类方法, 与公认的分类速度较快的 SEDS 几乎持平。未来的工作考虑将研究对象从一元时间序列延展到多元时间序列, 以构建有效的多元时间序列的 shapelets 学习模型。

参考文献

- [1] AU YEUNG J F K, WEI Z K, CHAN K Y, et al. Jump Detection in Financial Time Series Using Machine Learning Algorithms[J]. *Soft Computing*, 2020, 24(3): 1789-1801.
- [2] AL-HADEETHI H, ABDULLA S, DIYKH M, et al. Adaptive Boost LS-SVM Classification Approach for Time-Series Signal Classification in Epileptic Seizure Diagnosis Applications[J]. *Expert Systems with Applications*, 2020, 161: 113676.
- [3] ARUL M, KAREEM A. Applications of Shapelet Transform to Time Series Classification of Earthquake, Wind and Wave Data[J]. *Engineering Structures*, 2021, 228: 111564.
- [4] AHMED T, SINGH D. Probability Density Functions Based Classification of MODIS NDVI Time Series Data and Monitoring of Vegetation Growth Cycle[J]. *Advances in Space Research*, 2020, 66(4): 873-886.
- [5] 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.
- [6] RATANAMAHATANA C A, KEOGH E J. Three Myths about Dynamic Time Warping Data Mining[C]// *Proceedings of the 2005 SIAM International Conference on Data Mining*. 2005: 506-510.
- [7] DING H, TRAJCEVSKI G, SCHEUERMANN P, et al. Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures[J]. *Proceedings of the VLDB Endowment*, 2008, 1(2): 1542-1552.
- [8] 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.
- [9] KRUSKAL W H. A Nonparametric Test for the Several Sample Problem[J]. *Annals of Mathematical Statistics*, 1952, 23(4): 525-540.
- [10] THEOBALD R B C M. Introduction to the Theory of Statistics [J]. *Journal of the Royal Statistical Society Series A (General)*, 1974, 137(1): 95-96.
- [11] HILLS J, LINES J, BARANAUSKAS E, et al. Classification of Time Series by Shapelet Transformation[J]. *Data Mining and Knowledge Discovery*, 2014, 28(4): 851-881.
- [12] MUEEN A, KEOGH E, YOUNG N. Logical-Shapelets: An Expressive Primitive for Time Series Classification[C]// *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2011: 1154-1162.
- [13] 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.
- [14] YAHYAOU H, AL-DAIHANI R. A Novel Trend Based SAX Reduction Technique for Time Series[J]. *Expert Systems with Applications*, 2019, 130: 113-123.
- [15] PARK H, JUNG J Y. SAX-ARM: Deviant Event Pattern Discovery from Multivariate Time Series Using Symbolic Aggregate Approximation and Association Rule Mining[J]. *Expert Systems with Applications*, 2020, 141: 112950.
- [16] BUHLER J, TOMPA M. Finding Motifs Using Random Projections[J]. *Journal of Computational Biology*, 2002, 9(2): 225-242.
- [17] RAKTHANMANON T, KEOGH E. Fast shapelets: A Scalable Algorithm for Discovering Time Series Shapelets[C]// *Proceedings of the 2013 SIAM International Conference on Data Mining*. 2013: 668-676.

- [18] WISTUBA M, GRABOCKA J, SCHMIDT-THIEME L. Ultra-Fast Shapelets for Time Series Classification[J]. arXiv:1503.05018, 2015.
- [19] GORDON D, HENDLER D, ROKACH L. Fast and Space-Efficient Shapelets-Based Time-Series Classification[J]. Intelligent Data Analysis, 2015, 19(5):953-981.
- [20] LI Z S, HE Z F. Time Series Shapelet Extraction Based on Principal Component Analysis[J]. Computer Systems Applications, 2014, 23(11):145-149.
- [21] GRABOCKA J, SCHILLING N, WISTUBA M, et al. Learning Time-Series Shapelets [C] // Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014:392-401.
- [22] HOU L, KWOK J, ZURADA J. Efficient Learning of Timeseries Shapelets[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2016:1209-1215.
- [23] KARAMPATZIAKIS N, MINEIRO P. Discriminative Features Via Generalized Eigenvectors[C]//International Conference on Machine Learning, 2014:494-502.
- [24] TIBSHIRANI R, SAUNDERS M, ROSSET S, et al. Sparsity and Smoothness Via the Fused Lasso[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005, 67(1):91-108.
- [25] ZHANG Z, ZHANG H, WEN Y, et al. Discriminative Extraction of Features from Time Series[J]. Neurocomputing, 2018, 275:2317-2328.
- [26] SUGIYAMA M. Local Fisher Discriminant Analysis for Supervised Dimensionality Reduction[C]//Proceedings of the 23rd International Conference on Machine Learning, 2006:905-912.
- [27] WANG Y, EMONET R, FROMONT E, et al. Learning Interpretable Shapelets for Time Series Classification through Adversarial Regularization[J]. arXiv:1906.00917, 2019.
- [28] WANG Z, YAN W, OATES T. Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline[C]//2017 International Joint Conference on Neural Networks, 2017:1578-1585.
- [29] ZHAO H Y, PAN Z S. Multivariate Time Series Classification Based on Shapelets Learning [J]. Computer Science, 2018, 45(5):180-184, 219.
- [30] SIMON N, FRIEDMAN J, HASTIE T, et al. A Sparse-Group Lasso[J]. Journal of Computational and Graphical Statistics, 2013, 22(2):231-245.
- [31] TIBSHIRANI R. Regression Shrinkage and Selection Via the Lasso[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58(1):267-288.
- [32] YUAN M, LIN Y. Model Selection and Estimation in Regression with Grouped Variables[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2006, 68(1):49-67.
- [33] JI C, ZHAO C, LIU S, et al. A Fast Shapelet Selection Algorithm for Time Series Classification[J]. Computer Networks, 2019, 148:231-240.
- [34] ZHAO C, WANG T J, LIU S J, et al. A Fast Time Series Shapelet Discovery Algorithm Combining Selective Extraction and Subclass Clustering[J]. Journal of Software, 2020, 31(3):763-777.
- [35] FRIEDMAN J H. Regularized Discriminant Analysis[J]. Journal of the American Statistical Association, 1989, 84(405):165-175.



LI Chen, born in 1996, postgraduate. Her main research interests include data mining, machine learning and pattern recognition.



WAN Yuan, born in 1976, Ph.D, professor. Her main research interests include data mining, pattern recognition, manifold learning, machine learning and feature selection.

(责任编辑:杨雪敏)