基于全局−局部散度的多元时间序列无监督降维方法

李正欣[1,2]，胡钢[1]，张凤鸣[1]，张晓丰[1]，赵永梅[1]
（1. 空军工程大学装备管理与无人机工程学院，陕西 西安 710051；
2. 西北工业大学光电与智能研究院，陕西 西安 710072）

摘 要：针对传统降维方法不能直接应用于多元时间序列，现有的多元时间序列降维方法难以在保证降维有效性的同时大幅降低数据维度的问题，提出一种基于全局−局部散度的多元时间序列无监督降维方法。首先，提出一种特征序列提取方法，提取多元时间序列协方差矩阵的上三角元素，将其组合为特征序列。然后，以“局部散度最小、全局散度最大”为基本思想，提出一种无监督降维模型，在保持局部近邻关系的同时，尽可能保留全局信息。将特征序列作为输入，最小化所有样本点邻域方差之和，最大化邻域中心点方差。求解模型得到的投影矩阵能够实现多元时间序列的降维。最后，在 20 组公开数据集上，对所提方法进行了实验验证。结果表明，所提方法能够在保证降维有效性的同时，较大幅度地降低多元时间序列的维度。

关键词：多元时间序列；图结构；特征提取；无监督降维；分类精度
中图分类号：TP311
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2024008

Unsupervised dimensionality reduction method for multivariate time series based on global and local scatter

LI Zhengxin[1,2], HU Gang[1], ZHANG Fengming[1], ZHANG Xiaofeng[1], ZHAO Yongmei[1]
1. Equipment Management and Unmanned Aerial Vehicle Engineering School, Air Force Engineering University, Xi’an 710051, China
2. School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi’an 710072, China

Abstract: To solve the problem that the traditional dimensionality reduction methods cannot be directly applied to multivariate time series, and for the existing approaches, it is difficult to ensure the effectiveness of dimensionality reduction while significantly reducing the dimension, an unsupervised dimensionality reduction method of multivariate time series based on global and local scatter was proposed. Firstly, a feature series extraction method was proposed to extract the upper triangular elements of the co-variance matrix of each multivariate time series and combine them into a feature sequence. Then, based on the idea of “minimum local scatter and maximum global scatter,” an unsupervised dimensionality reduction model was presented, which preserved the global information as much as possible while maintaining the local nearest neighbor relationship. Using the feature sequence as the input, the sum of the neighborhood variances of all sample points was minimized, and the variance of all the neighborhood centroids were maximized. The projection matrix obtained by solving the proposed model could be used to perform the dimensionality reduction. Finally, the proposed method was evaluated with experiments on 20 public data sets. The results show that the proposed method can significantly reduce the dimension of multivariate time series, while ensuring the effectiveness of dimensionality reduction.

Keywords: multivariate time series, graph structure, feature extraction, unsupervised dimensionality reduction, classification accuracy

收稿日期：2023−04−12；**修回日期：**2023−07−04
基金项目：国家自然科学基金资助项目（No.62002381）
Foundation Item: The National Natural Science Foundation of China (No.62002381)

0 引言

时间序列是一种常见的数据类型，广泛存在于金融[1]、航空航天[2]、医学[3]、环境[4]、交通[5]、气象[6]等领域，如“黑匣子”记录的飞行数据、大脑产生的脑电信号、物体运动产生的加速度等。此外，物体的轮廓数据、视频中提取的视频帧[7]也可以被视为时间序列。时间序列蕴含系统内部变化规律，利用它可以了解系统发展趋势。近年来，时间序列数据挖掘已成为热门研究领域之一，研究内容主要包括分类、聚类、预测、相似性搜索、模式匹配等。

时间序列是按时间排序组成的数据集合[8]。数值型时间序列可表示为 $\boldsymbol{X}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)^{\mathrm{T}}\in\mathbb{R}^{m\times t}$，$m$ 为变量数，$m=1$ 时，$\boldsymbol{X}$ 为一元时间序列（UTS, univariate time series），$m>1$ 时，$\boldsymbol{X}$ 为多元时间序列（MTS, multivariate time series）；t 为时间长度。现实应用中，系统往往需要多个变量进行描述，因此 MTS 比 UTS 更常见。随着数据采集技术的发展，各领域产生的 MTS 维度显著增加、规模不断扩大，给后续的分析处理带来了较大挑战。“维度灾难”不仅导致计算效率较低，还会造成精度下降[9]。因此，在进行数据挖掘之前，通常需要对时间序列进行降维。

降维是处理高维数据的一个必要环节，其目的是在保留数据本质特征的前提下，降低数据维度[10]。在 MTS 数据集中，一个样本序列可以表示为矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times t}$。MTS 的降维任务是将高维矩阵 $\boldsymbol{X}$ 映射为低维矩阵 $\boldsymbol{Y}\in\mathbb{R}^{l\times h}$，使 $l\leqslant m$ 且 $h\leqslant t$。降维技术通常分为特征选择和特征提取。特征选择是从原始特征中选择一个特征子集。特征提取通过映射从原始特征中得到新的特征。特征提取的优点是只需要更少的新特征便能反映原始数据，使数据的压缩效率更高[11]。因此，本文主要研究 MTS 特征提取方法。

MTS 同时具有时间维度和变量维度，且样本的时间长度不一定相等。因此，传统的特征提取方法不能很好地适用于 MTS。目前，MTS 特征提取方法主要包括主成分分析（PCA, principal component analysis）[12-14]、奇异值分解（SVD, singular value decomposition）[15]、独立成分分析（ICA, independent component analysis）[16]等。通常，这些方法对数据集中的 MTS 逐个进行降维，忽略了不同 MTS 之间的关系。

近年来，基于图结构的特征提取方法逐渐成为研究热点，在常规高维数据的模式识别领域取得了较好效果。原因在于基于图结构的特征提取通过构建图结构能够有效保留高维空间样本点之间的近邻关系。因此，基于图结构的方法也被应用到 MTS 特征提取[17-23]，具体包括以下 2 种研究思路。

一种思路是直接对原始序列建立图结构进行降维。文献[17]从行和列 2 个方向对 MTS 同时使用线性判别分析（LDA, linear discriminant analysis）实现降维。文献[18]对等长 MTS 数据集采用拉普拉斯映射（LE, Laplacian eigenmaps）实现非线性降维。对原始序列直接建立图结构进行降维通常要求数据集中 MTS 的时间维度相等，通过计算不同 MTS 之间的相似度建立图结构进行降维。在处理不等长 MTS 数据集时，通常会采用截断或者补全的方法先将不等长 MTS 变成等长 MTS，这个过程可能会造成信息丢失或信息冗余。

另一种思路是采用“两阶段”降维方法，第一阶段是将 MTS 数据集转换为等长特征序列集，第二阶段是对特征序列建立图结构进行降维。文献[19-21]在第一阶段均使用 SVD 提取 MTS 的奇异向量作为特征序列，在第二阶段使用局部保持投影（LPP, locality preserving projection）或其改进算法对特征序列进行投影，实现降维。文献[22]提取 MTS 的时域和频域特征作为特征序列，之后利用改进的局部线性嵌入（M-LLE, modified local linear embedding）对特征序列进行降维。文献[23]计算 MTS 协方差矩阵作为特征序列，之后利用标签值构建二维类间边界 Fisher 分析模型实现对特征序列的降维。MTS 通常维度高、信息密度低，“两阶段”降维方法的优势在于降维幅度较大，能够极大地消除 MTS 的冗余信息，且能较好地保留特征序列之间的近邻关系。但是，“两阶段”降维方法的性能很大程度取决于所提取特征序列的质量。现有的“两阶段”降维方法的主要问题在于提取特征序列的有效性不足，图结构没有同时考虑全局和局部信息，导致该类方法难以实现降维幅度和降维效果的有效平衡。如何有效提取等长特征序列、如何合理构建图结构成为解决以上问题的关键。为此，本文首先提取 MTS 协方差矩阵的上三角元素，将其组合为特征序列。然后，以“局部散度最小、全局散度最大”为基本思想，提出一种基于图结构的特征提取方法，将特征序列投影到低维空间实现降维。所提方法能够在保留 MTS 有效信息的基础上，实现较大幅度

的降维。最后，通过实验验证了所提方法的有效性。本文的研究工作主要如下。

1) 提出一种 MTS 特征序列提取方法。计算 MTS 的协方差矩阵，提取协方差矩阵的上三角元素，将其组合为特征序列。既保留了变量间的相关信息，又能够将数据集中所有等长或不等长 MTS 转化为等长的特征序列。

2) 以“局部散度最小、全局散度最大”为基本思想，提出一种基于全局−局部散度的无监督降维方法，简称 GLSUP 方法，在保持局部近邻关系的同时，尽可能地保留全局信息。

3) 使用公开数据集对所提方法进行了实验验证。实验结果表明，所提方法不仅具有较好的降维有效性，而且能够较大幅度地降低 MTS 维度。

1　相关工作

传统基于 PCA 的 MTS 降维方法是对每个 MTS 单独进行 PCA 降维，根据方差贡献率确定降维幅度[24]。该方法计算简单，考虑了 MTS 变量间的相关性，但它没有将所有 MTS 投影到同一个低维子空间，使不同 MTS 降维后，各个变量含义不同。为此，文献[25]提出基于共同主成分分析（CPCA, common principal component analysis）的 MTS 降维方法，将所有 MTS 投影到同一个子空间，解决了传统 PCA 在 MTS 降维时存在的问题。此外，文献[26]提出一种基于 CPCA 的 MTS 投影方法，称为 CCPCA 方法，它利用数据标签对相同类别的 MTS 进行 CPCA 投影，形成多个不同的公共投影子空间。CCPCA 方法适用于类别投影子空间差异较大的 MTS 数据集，当数据集类别差异不大时，CCPCA 方法与 CPCA 方法效果相近。然而，CPCA 方法和 CCPCA 方法都只能表示变量之间的线性关系。为了实现非线性降维，文献[27]提出基于共同核主成分分析（CKPCA）的 MTS 降维方法。

文献[12]提出基于变量的 PCA（VPCA, variable-based PCA）降维方法。该方法将所有 MTS 的同一变量对应的序列组合成一个矩阵，对每个组合后的矩阵进行 PCA 降维。其本质是对 MTS 的时间维度降维，优点在于考虑了不同 MTS 同一变量之间的信息，但没有考虑变量间的相关性，且只适用于等长 MTS 数据集。文献[28]提出基于频率主成分的降维方法 FC_PCA。该方法提出一种新的谱域方法，将多元二阶平稳时间序列表示为多个频率上的子序列，之后在特定频带上进行 PCA 降维。FC_PCA 方法在某些类型的数据上能够获得较好的效果，但要求时间序列为二阶平稳序列。文献[29]提出基于变量相关性的降维方法。该方法计算 MTS 的协方差矩阵，将多个协方差矩阵拼接起来，而后进行主元分析，得到 MTS 的低维特征矩阵，能够处理不等长 MTS 数据集，但得到的特征矩阵不能确保在同一低维子空间。文献[14]提出一种基于 PCA 分段表示（PPCA, piecewise representation based on PCA）的降维方法。该方法对每个 MTS 进行分段并分别计算协方差，通过 PCA 提取得到平均协方差矩阵的主成分，从而实现降维。该方法考虑了 MTS 变量之间的相关关系，但是没有将 MTS 投影到同一个低维子空间，投影后不同 MTS 的各变量含义不同。

以上基于 PCA 的降维方法只考虑了样本点的全局方差，而忽略了样本点的局部信息。在常规高维数据降维中，基于图结构的方法能够有效体现样本点的局部信息，常见方法包括局部保持投影[30]、边界 Fisher 分析（MFA, marginal Fisher analysis）[7, 31]、局部线性嵌入（LLE, locally linear embedding）[32]等。文献[33]提出了一种用于常规高维数据的无监督投影算法 GLUP（globally and locally consistent unsupervised projection），同时考虑全局和局部一致性。该方法借鉴了 PCA 方法的“方差最大化”思想，通过邻域方差和全局方差分别建立局部结构模型和全局结构模型。

在 MTS 降维中，学者对基于图结构的降维方法进行了研究。当 MTS 数据集中的序列均等长时，可以直接对原始序列建立图结构进行降维。PBLDA（pseudo bidirectional LDA）[17]从 MTS 行和列 2 个方向分别执行线性判别分析，得到 2 个方向的投影矩阵，实现降维。PBLDA 能够同时缩减变量维度和时间维度，但是只能用于等长数据集，处理不等长数据集时要采用截断或者补齐的方法将不等长 MTS 变为等长 MTS，存在丢失信息风险。文献[18]将 LE 用于 MTS 降维，首先针对大脑不同状态采集得到的功能核磁共振成像数据构建相似性图，之后计算图的拉普拉斯矩阵并通过特征分解得到降维特征序列。LE 属于非线性降维方法，能够较好地处理非线性数据，但是其对噪声比较敏感，对于含噪声的数据可能无法建立较可靠的图结构。

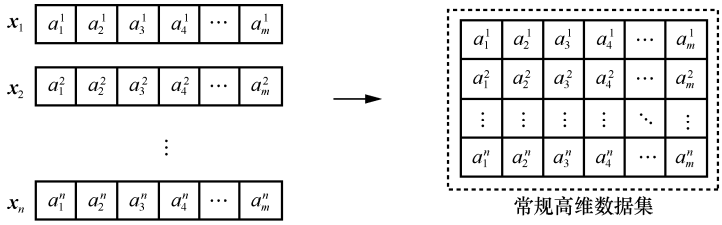
当处理的是不等长数据集时，直接对原始序列降维的方法通常需要对原始序列进行截断或者补全，而“两阶段”降维方法可以更好地处理不等长数据集，获得更大的降维幅度。SVD_LPP[19]采用基于 SVD 的方法从 MTS 中提取特征序列，之后采用 LPP 方法对特征序列进行降维。为了利用标签值，文献[20]提出了基于有监督局部保持投影的降维方法，在近邻选择中引入了标签值，但是没有考虑数据样本的类内和类间离散程度。为此，文献[21]提出了基于奇异值分解和判别局部保持投影（S_DLPP, SVD and discriminant LPP）的降维方法。该方法能够从局部结构上实现投影后“同类近邻点近、异类近邻点远”，但是没有考虑数据集的全局结构信息。M-LLE（modified LLE）[22]首先提取 MTS 的时域和频域特征组成特征序列，之后采用马氏距离（MD）度量代替传统邻域构造中的欧氏距离（ED, Euclid distance），并采用 L1 范数规范权值矩阵，通过 LLE 实现对特征序列的降维。M-LLE 通过 L1 范数增强了算法的抗噪能力，但是没有考虑数据的全局结构信息。文献[34]提出了基于特征选择的 Shapelets 发现（LSDF）方法，将 MFA 和融合 Lasso 算子结合得到判别特征 Shapelets。但 LSDF 方法仅针对 UTS，且寻找 Shapelets 的时间代价较高。文献[23]提出了基于二维类间边界 Fisher 分析（2DICMFA, two-dimensional inter-class marginal Fisher analysis）的 MTS 降维方法，针对边界 Fisher 分析进行模型改进，提出类间边界 Fisher 分析模型并进行二维化拓展，得到二维类间边界 Fisher 分析的降维模型，通过计算协方差矩阵将多元时间序列转化为特征矩阵，利用降维模型将特征矩阵投影到一个低维空间，实现特征降维。2DICMFA 方法克服了 MFA 无法处理不同类别样本边界不明确的问题，但是需要样本标签值，且采用的特征矩阵为对称矩阵，存在冗余信息。

当前“两阶段”降维方法主要问题如下：提取特征序列的有效性不足，采用的图结构没有同时考虑全局和局部信息，导致该类方法难以实现降维幅度和降维效果的有效平衡。针对以上问题，本文首先提取 MTS 协方差矩阵的上三角元素，将其组合为特征序列；然后，提出一种基于全局−局部散度的无监督降维方法。在保证降维有效性的同时，较大幅度地降低 MTS 的维度。

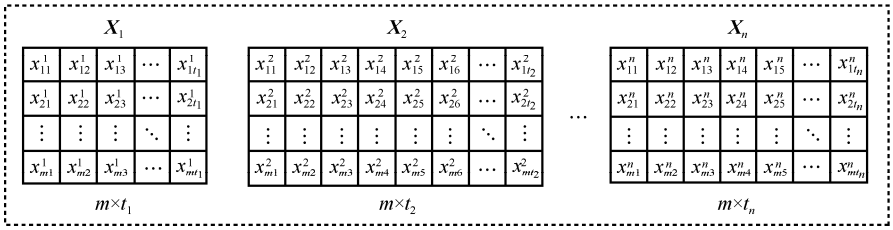
2 MTS 特征序列提取

与常规高维数据相比，MTS 高维特性如图 1 所示。常规高维数据集的一个样本可以用 $1\times m$ 的向量表示，且每个样本的维度相同。因此，常规高维数据集可以用一个矩阵表示。MTS 具有 2 种维度属性：时间维和变量维。一个 MTS 用 $m\times t_i$ 的矩阵表示。通常，在一个 MTS 数据集中，不同样本的变量维度相同、时间长度不一定相同。因此，MTS 数据集需要用多个长度不同的矩阵表示。基于上述原因，传统降维方法并不能直接应用于 MTS。

设 MTS 数据集 $D=\{\boldsymbol{X}_i \mid i=1,2,\cdots,n\}$，其中，$n$ 为样本个数，$\boldsymbol{X}_i=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m)^{\mathrm{T}}\in\mathbb{R}^{m\times t_i}$ 为数据集



(a) 常规高维数据表示方法



(b) MTS高维数据表示方法

图 1 常规高维数据和 MTS 高维数据表示方法

中第 i 个 MTS 样本，$\boldsymbol{x}_i(i=1,2,\cdots,m)$ 为第 i 个变量的一系列观测值，m 为变量数，t_i 为第 i 个 MTS 的时间长度，不同 MTS 的时间长度可能不等。为此，本文提出一种 MTS 特征序列提取方法。首先，对每个 MTS 进行零均值化处理，即 $\boldsymbol{X}_i=\boldsymbol{X}_i-\mathrm{E}(\boldsymbol{X}_i)$。此时，第 i 个 MTS 的协方差矩阵为

$$\boldsymbol{\sigma}_i=\mathrm{E}(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}})=\begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix} \tag{1}$$

MTS 的协方差矩阵 $\boldsymbol{\sigma}_i\in\mathbb{R}^{m\times m}$ 为对称阵，将 $\boldsymbol{\sigma}_i$ 的上三角元素提取出来，按顺序组成行向量，表示为

$$\boldsymbol{f}_i=(a_{11},\cdots,a_{1m},a_{21},\cdots,a_{2m},a_{m1},\cdots,a_{mm}) \tag{2}$$

将 $\boldsymbol{f}_i$ 作为第 i 个 MTS 的特征序列，得到 MTS 的特征集 $\mathrm{Fea}=\{\boldsymbol{f}_i\,|\,i=1,2,\cdots,n\}$，MTS 特征序列提取过程如图 2 所示。从图 2 可以看出，经过上述特征提取过程后特征序列 $\boldsymbol{f}_i$ 的长度为 $\dfrac{m(m+1)}{2}$，与时间长度 t_i 无关。因此，通过该方法不仅可以提取 MTS 不同变量之间的相关性特征，并且得到的特征序列是等长的。

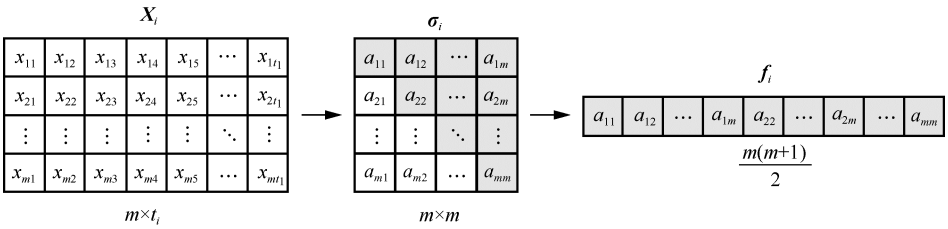


图 2　MTS 特征序列提取过程

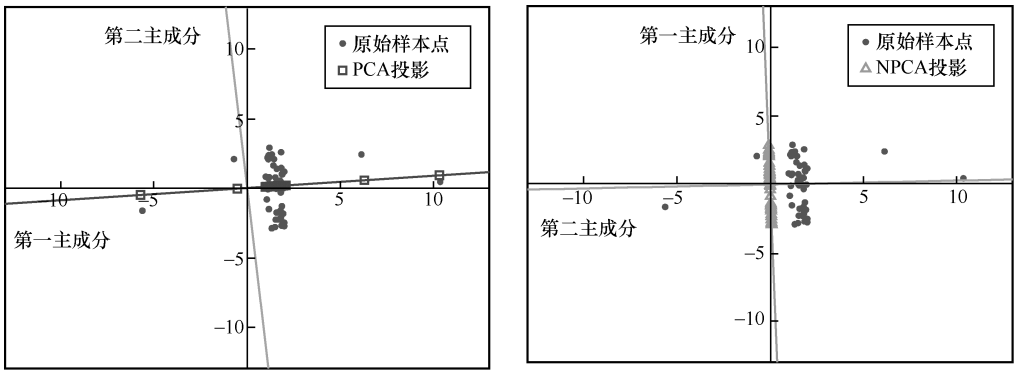
3　基于全局−局部散度的无监督降维方法

3.1　基本思想

图结构中的“局部散度”表示数据集局部离散程度，“局部散度最小”使高维空间中的近邻样本点在投影后依旧保持近邻关系。“全局散度”表示数据集的全局离散程度，“全局散度最大”可使投影后的样本点保留更多的全局信息。本节基于“局部散度最小、全局散度最大”思想，提出一种无监督降维模型。

“局部散度”通过样本点的邻域方差来体现，这种方法相比于 LPP 方法需要的参数更少。“全局散度”通过本文提出的无监督邻域 PCA（NPCA, neighbor PCA）方法体现。与 PCA 方法相比，NPCA 方法能够降低离群点的影响。PCA 投影和 NPCA 投影如图 3 所示。原始样本点中有 46 个高斯分布的正常样本点以及 4 个离群点。从图 3(a)可以看出，由于离群点的影响，正常的原始样本点通过 PCA 方法投影后的方差并没有最大化，反而变得更密集，不利于全局信息的保留。

首先，计算每个原始样本点的邻域点集合；然后，针对原始样本点及其邻域点组成的点集，计算邻域中心点；最后，用邻域中心点替代原始样本点来计算最优投影方向，以保留最大全局方差。在图 3(b)中，NPCA 方法的近邻数 k 取 10，得到最优投影方向。可以看出，正常样本点投影后，能够保留最大方差，说明 NPCA 方法能够克服离群点的影响。



(a) PCA投影　　(b) NPCA投影

图 3　PCA 投影和 NPCA 投影

3.2 模型建立

通过第 2 节的特征序列提取，从 MTS 数据集 D 中得到特征集 $\text{Fea}=\{\boldsymbol{f}_i \mid i=1,2,\cdots,n\}$，$\boldsymbol{f}_i \in \mathbb{R}^p$，其中，$p=\dfrac{m(m+1)}{2}$，$m$ 是 MTS 变量数。降维特征集 $D'=\{\boldsymbol{y}_i \mid i=1,2,\cdots,n\}$，$\boldsymbol{y}_i=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{f}_i \in \mathbb{R}^d$（$d \ll p$）。首先，使用 k 近邻和欧氏距离度量针对特征集建立邻域集合 $N_k(\boldsymbol{f}_i)=\left\{\boldsymbol{f}_j \mid j=1,2,\cdots,k\right\}$，其中 $\boldsymbol{f}_j$ 为 $\boldsymbol{f}_i$ 的 k 近邻点。在找到每个样本点的邻域后，计算特征集中每个特征序列 $\boldsymbol{f}_i$ 的邻域中心序列 $\boldsymbol{m}_i$，即

$$\boldsymbol{m}_i=\frac{1}{k+1}\left(\sum_{\boldsymbol{f}_j \in N_k(\boldsymbol{f}_i)} \boldsymbol{f}_j+\boldsymbol{f}_i\right) \tag{3}$$

其中，k 为近邻数，$\boldsymbol{m}_i$ 为 $\boldsymbol{f}_i$ 与其 k 个近邻点的邻域中心序列。

3.2.1 局部散度

用投影后样本点的邻域方差表征局部散度。先计算投影后每个样本点邻域集合的方差，再对这些方差累加求和，可得

$$J_{\mathrm{L}}(\boldsymbol{W})=\frac{1}{2}\sum_{i=1}^{n}\sum_{\boldsymbol{f}_j \in N_k(\boldsymbol{f}_i)}(\boldsymbol{y}_j-\boldsymbol{p}_i)^2 \tag{4}$$

其中，$\boldsymbol{p}_i=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{m}_i$ 为邻域中心点的低维投影。对式(4)进行变换可得

$$\begin{aligned}&J_{\mathrm{L}}(\boldsymbol{W})=\frac{1}{2}\sum_{i=1}^{n}\sum_{\boldsymbol{f}_j \in N_k(\boldsymbol{f}_i)\cup \boldsymbol{f}_i}(\boldsymbol{y}_j-\boldsymbol{p}_i)^2=\\&\frac{1}{2}\sum_{i=1}^{n}\sum_{\boldsymbol{f}_j \in N_k(\boldsymbol{f}_i)\cup \boldsymbol{f}_i}(\boldsymbol{W}^{T}\boldsymbol{f}_j-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{m}_i)^2=\\&\boldsymbol{W}^{\mathrm{T}}\left[\frac{1}{2}\sum_{i=1}^{n}\sum_{\boldsymbol{f}_j \in N_k(\boldsymbol{f}_i)\cup \boldsymbol{f}_i}(\boldsymbol{f}_j-\boldsymbol{m}_i)^2\right]\boldsymbol{W}=\\&\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{L}}\boldsymbol{W}\end{aligned} \tag{5}$$

其中，$\boldsymbol{S}_{\mathrm{L}}=\sum_{i=1}^{n}\boldsymbol{S}_{\mathrm{L}_i}$，$\boldsymbol{S}_{\mathrm{L}_i}$ 为第 i 个特征序列 $\boldsymbol{f}_i$ 的局部散度矩阵，即

$$\boldsymbol{S}_{\mathrm{L}_i}=\sum_{\boldsymbol{f}_j \in N_k(\boldsymbol{f}_i)\cup \boldsymbol{f}_i}(\boldsymbol{f}_j-\boldsymbol{m}_i)(\boldsymbol{f}_j-\boldsymbol{m}_i)^{\mathrm{T}} \tag{6}$$

3.2.2 全局散度

为了克服离群点的影响，NPCA 方法使用邻域中心点来计算全局方差。则全局散度定义为

$$J_{\mathrm{G}}(\boldsymbol{W})=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{p}_i-\boldsymbol{p}_j)^2 \tag{7}$$

其中，$\boldsymbol{p}_i=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{m}_i$ 为邻域中心点的低维投影。对式(7)进行变换可得

$$\begin{aligned}&J_{\mathrm{G}}(\boldsymbol{W})=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{p}_i-\boldsymbol{p}_j)^2=\\&\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{m}_i-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{m}_j)^2=\\&\boldsymbol{W}^{\mathrm{T}}\left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{m}_i-\boldsymbol{m}_j)(\boldsymbol{m}_i-\boldsymbol{m}_j)^{\mathrm{T}}\right]\boldsymbol{W}=\\&\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{G}}\boldsymbol{W}\end{aligned} \tag{8}$$

其中，$\boldsymbol{S}_{\mathrm{G}}$ 为全局散度矩阵，表达式为

$$\boldsymbol{S}_{\mathrm{G}}=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{m}_i-\boldsymbol{m}_j)(\boldsymbol{m}_i-\boldsymbol{m}_j)^{\mathrm{T}} \tag{9}$$

对其进行化简可得

$$\begin{aligned}&\boldsymbol{S}_{\mathrm{G}}=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{m}_i-\boldsymbol{m}_j)(\boldsymbol{m}_i-\boldsymbol{m}_j)^{\mathrm{T}}=\\&\frac{1}{2}\left(\sum_{i=1}^{n}\boldsymbol{m}_i\boldsymbol{m}_i^{\mathrm{T}}+\sum_{j=1}^{n}\boldsymbol{m}_j\boldsymbol{m}_j^{\mathrm{T}}-2\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{m}_i\boldsymbol{m}_j^{\mathrm{T}}\right)=\\&\left(\sum_{i=1}^{n}\boldsymbol{m}_i\boldsymbol{m}_i^{\mathrm{T}}-\sum_{i=1}^{n}\boldsymbol{m}_i\sum_{j=1}^{n}\boldsymbol{m}_j^{\mathrm{T}}\right)=\sum_{i=1}^{n}(\boldsymbol{m}_i-\boldsymbol{m}_0)(\boldsymbol{m}_i-\boldsymbol{m}_0)^{\mathrm{T}}\end{aligned} \tag{10}$$

其中，$\boldsymbol{m}_0=\dfrac{1}{n}\sum_{j=1}^{n}\boldsymbol{m}_j$ 为所有邻域中心的中心点，即全局邻域中心。

3.2.3 模型构建

为了使高维空间中的近邻样本点在投影后依旧保持近邻关系，同时保留更多的全局信息，以“局部散度最小、全局散度最大”为基本思想，可得 GLSUP 模型为

$$\max_{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{I}} J_1(\boldsymbol{W})=\mathrm{Tr}\left(\frac{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{G}}\boldsymbol{W}}{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{L}}\boldsymbol{W}}\right) \tag{11}$$

3.3 模型求解

根据文献[30]，式(11)可转化为广义特征值求解问题，即

$$\boldsymbol{S}_{\mathrm{G}}\boldsymbol{\omega}=\lambda\boldsymbol{S}_{\mathrm{L}}\boldsymbol{\omega} \tag{12}$$

其中，$\boldsymbol{\omega}$ 为广义特征向量，λ 为广义特征值。通过求解式(12)，可以得到前 d 个最大的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_d(d<p)$ 以及对应的特征向量 $\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\cdots,\boldsymbol{\omega}_d$，投影矩阵 $\boldsymbol{W}=(\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\cdots,\boldsymbol{\omega}_d)$。根据投影矩阵，计算降维特征序列为

$$\boldsymbol{y}_i = \boldsymbol{W}^{\mathrm{T}} \boldsymbol{f}_i \tag{13}$$

最终，得到降维特征集 $D' = \{\boldsymbol{y}_i \,|\, i = 1,2,\cdots,n\}$，其中，$\boldsymbol{y}_i \in \mathbb{R}^d$。GLSUP 方法的计算步骤如算法 1 所示。首先，从 MTS 的协方差矩阵中提取得到特征序列，组成特征集 Fea；然后，根据特征序列计算局部散度矩阵 $\boldsymbol{S}_{\mathrm{L}}$ 和全局散度矩阵 $\boldsymbol{S}_{\mathrm{G}}$；最后，通过模型求解得到降维特征集 D'。

算法 1　GLSUP 方法

输入　MTS 数据集 D，近邻数 k

输出　降维特征集 D'

1) for $i \leftarrow 1$ to n do
2)　$\boldsymbol{X}_i = \boldsymbol{X}_i - \mathrm{mean}(\boldsymbol{X}_i)$; //数据零均值化
3)　$\boldsymbol{\sigma}_i = \boldsymbol{X}_i \boldsymbol{X}_i^{\mathrm{T}}$; //计算协方差矩阵
4)　$\boldsymbol{f}_i = \mathrm{ExtrcatFea}(\boldsymbol{\sigma}_i)$; //提取特征序列
5)　$N_k(\boldsymbol{f}_i) = \mathrm{knnsearch}(\boldsymbol{f}_i, k)$; //寻找样本邻域
6)　$\boldsymbol{m}_i = \mathrm{mean}(N_k(\boldsymbol{f}_i) + \boldsymbol{f}_i)$ //计算邻域中心
7) end for
8) $\boldsymbol{m}_0 = \mathrm{mean}(\boldsymbol{m}_i)$; //计算全局邻域中心 $\boldsymbol{m}_0$
9) $\boldsymbol{S}_{\mathrm{L}} = \mathrm{zeros}(p, p)$; //局部散度初始化
10) $\boldsymbol{S}_{\mathrm{G}} = \mathrm{zeros}(p, p)$; //全局散度初始化
11) for $i \leftarrow 1$ to n do
12)　for $j \leftarrow 1$ to $k+1$ do
13)　$\boldsymbol{S}_{\mathrm{L}} = \boldsymbol{S}_{\mathrm{L}} + (\boldsymbol{f}_j - m_i)(\boldsymbol{f}_j - m_i)^{\mathrm{T}}$; //计算局部散度
14)　end for
15)　$\boldsymbol{S}_{\mathrm{G}} = \boldsymbol{S}_{\mathrm{G}} + (\boldsymbol{m}_i - \boldsymbol{m}_0)(\boldsymbol{m}_i - \boldsymbol{m}_0)^{\mathrm{T}}$; //计算全局散度
16) end for
17) $W = \mathrm{eigs}(\boldsymbol{S}_{\mathrm{L}}, \boldsymbol{S}_{\mathrm{G}})$; //计算投影矩阵
18) for $i \leftarrow 1$ to n do
19)　$\boldsymbol{y}_i = W^{\mathrm{T}} \boldsymbol{f}_i$; //计算降维特征序列
20)　$D'(i) = \boldsymbol{y}_i$; //得到降维特征集
21) end for

3.4　计算复杂度分析

设 MTS 数据集 $D = \{\boldsymbol{X}_i \,|\, i = 1,2,\cdots,n\}$，其中，$n$ 为样本个数，$\boldsymbol{X}_i = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m)^{\mathrm{T}} \in \mathbb{R}^{m \times t}$ 为数据集中第 i 个 MTS，$\boldsymbol{x}_i (i = 1,2,\cdots,m)$ 为第 i 个变量的一系列观测值，m 为变量数，特征序列长度为 $\dfrac{m(m+1)}{2}$。假设数据集中 MTS 的平均长度为 t，降维之后的维度为 d（$d \ll m^2$）。

GLSUP 方法的计算复杂度主要包括三部分：获取特征序列、模型求解以及投影。获取特征序列的计算代价主要是 MTS 协方差矩阵的计算，计算复杂度为 $O(nm^2t)$。模型求解的计算代价主要是局部散度矩阵和广义特征值求解的计算，计算复杂度分别为 $O(nm^4)$ 和 $O(m^6)$。在求得投影矩阵 $\boldsymbol{W}$ 后，GLSUP 方法对特征集 Fea 中的特征序列进行投影，计算复杂度为 $O(dm^2)$。因此，GLSUP 方法的复杂度为 $O(nm^2t + nm^4 + m^6)$。

4　实验分析

4.1　实验环境与数据

硬件环境如下：Intel(R) Core(TM)i7-9750HF CPU，256 GB+1 TB 硬盘，32 GB 内存。软件环境如下：Windows 10 专业版，MATLAB 2020b。实验采用 20 个应用广泛的公开数据集[35-36]，分别为 LP1、LP2、LP3、LP4、LP5、DSA（daily and sports activities）、FM（finger movements）、HMD（hand movement direction）、NATOPS（naval air training and operating procedures standardization）、Cricket、RS（racket sports）、Epilepsy、BM（basic motions）、LSST（large synoptic survey telescope）、AWR（articulary word recognition）、EEGeye、Wafer、WR（WalkvsRun）、KP（Kickvs Punch）、ASL（Australian sign language）。MTS 数据集信息如表 1 所示，前 15 个为等长数据集，后 5 个为不等长数据集。

表 1　MTS 数据集信息

编号	数据集	类别	特征维度	序列长度	序列均长	样本数
1	LP1	4	6	15	15	88
2	LP2	5	6	15	15	47
3	LP3	4	6	15	15	47
4	LP4	5	6	15	15	47
5	LP5	4	6	15	15	47
6	DSA	19	45	125	125	2 400
7	FM	2	28	50	50	416
8	HMD	4	10	400	400	234
9	NATOPS	6	24	51	51	360
10	Cricket	12	6	1197	1197	180
11	RS	4	6	30	30	303
12	Epilepsy	4	3	206	206	275
13	BM	4	6	100	100	80
14	LSST	14	6	36	36	4 925
15	AWR	25	9	144	144	575
16	EEGeye	2	14	20~2 401	624	24
17	Wafer	2	6	104~198	137	1 194
18	WR	2	62	128~1 918	368	44
19	KP	2	62	274~841	427	26
20	ASL	95	22	45~136	57	2 565

Robot Execution Failure 为机器人故障监控数据，包括 5 个子集，即 LP1~LP5。DSA 记录了 8 名实验者进行 19 项活动时产生的体感数据，实验选用前 5 个活动的数据集。FM 记录了食指和小指键入符号时，产生的脑电信号。HMD 记录了实验者在听到提示后，使用手腕进行 4 个不同方向运动时产生的脑电信号。NATOPS 记录了美国海军舰载机引导员进行指挥时，手、肘、腕和拇指上共 24 个传感器产生的加速度数据。Cricket 记录了板球裁判员进行手语裁判时，手腕加速度数据。RS 是智能手表采集得到的测试者在进行体育运动时的坐标和加速度数据。Epilepsy 记录了测试者在模拟课堂上进行缓慢行走、边做手势边缓慢行走、正常行走和快速行走等 4 种不同活动时，手腕加速度数据。BM 记录了测试者穿戴智能手表进行站立、行走、跑步和打羽毛球等 4 种活动时的 3D 加速度计和 3D 陀螺仪数据。LSST 是通过 6 个不同天文滤波器中的光子通量，记录物体亮度的天文时间序列数据，天文滤波器的光谱包括紫外光、可见光和红外光区域。AWR 记录了人发出不同单词时，嘴唇和舌头运动的传感器数据。

EEGeye 记录了人睁眼、闭眼时的脑电数据。Wafer 用 6 个真空传感器，记录了硅晶体产生时的半导体微电子序列。WR 记录了人行走或奔跑时的姿态数据，不同样本的时间长度差异较大。KP 记录了测试者脚踢和拳击时，身体加速度数据。ASL 为澳大利亚手语数据集，包含 95 种手语含义，共有 22 个变量，其中，左、右手各用 11 个变量刻画动作特征：5 个变量表示五根手指的弯曲程度，6 个变量表示手所处的位置。

4.2 实验方法与参数设置

降维有效性是指数据降维后，保留信息对 MTS 特征的刻画程度。实验将降维结果输入 KNN 分类器（K=1）中，通过分类精度来评估降维有效性。使用降维算法对原始 MTS 数据集进行降维，得到降维数据集。从降维数据集中依次选取样本输入分类器中，使用最近邻查询得到一个与被查询样本最相似的样本，将该样本标签值作为被查询样本所属类别，若与被查询样本标签值一致，则为正确分类，否则为不正确分类。在对所有样本执行完操作后，得到分类精度 ε 为

$$\varepsilon = \frac{n_{\text{true}}}{n} \tag{14}$$

其中，n_{true} 为正确分类的样本数量，n 为样本个数。

选取 PCA、CPCA、PPCA、SVD_LPP、PBLDA、VPCA 等 6 种降维方法作为对比方法。由于 VPCA 方法只适用于等长数据集，只在 15 组等长数据集上实验。在不等长数据集上，PCA、CPCA 和 PPCA 方法的降维结果仍为不等长序列。由于欧氏距离只能度量等长序列，在 5 组不等长数据集中，使用动态时间弯曲（DTW, dynamic time warping）距离对 PCA、CPCA 和 PPCA 方法的降维结果进行度量，实现 KNN 分类。

除了降维有效性，降维幅度也是降维算法的重要衡量指标。降维幅度是指减少的数据维度。实验将降维幅度定义为

$$\text{comp} = \frac{m\bar{t} - d\bar{t}'}{m\bar{t}} \tag{15}$$

其中，comp 为降维幅度，m 为原始 MTS 的变量数，$\bar{t}$ 为原始 MTS 的平均序列长度，d 为降维后特征数据的变量数，$\bar{t}'$ 为降维后，特征序列的平均长度。

PCA、CPCA、VPCA 方法涉及的参数是方差贡献率 σ。PPCA 方法涉及的参数有方差贡献率 σ、分段数量 ω。SVD_LPP 方法涉及的参数有近邻数 k、热核参数 t、降维之后的维度 d。PBLDA 方法涉及的参数有特征维度 p_c、时间维度 p_r。GLSUP 方法涉及的参数有近邻数 k、降维之后的维度 d。实验中，将 PCA、CPCA、PPCA 和 VPCA 方法的方差贡献率 σ 设为 80%，近邻数 k 和热核参数 t 均设为 1，时间维度 p_r 与原时间序列长度保持不变。PPCA 的最优匹配精度通过对分段数量 ω 从[2, 5, 10]中取值来获得。PBLDA、GLSUP、SVD_LPP 的最优分类精度通过调整特征维度 p_c 和降维之后的维度 d 获得。

4.3 降维有效性分析

降维有效性实验结果如表 2 所示，表 2 中每行的最优分类精度用粗体表示，次优分类精度用下划线表示。从结果来看，GLSUP 方法在 9 个数据集中取得了最优分类精度，在 8 个数据集中取得了次优的分类精度。GLSUP 方法将 MTS 转化为等长特征序列，保留了不同变量之间的相关性信息，之后同时考虑了数据集的全局和局部信息，对特征序列实现了降维。

表 2　　**降维有效性实验结果**

数据集编号	GLSUP	PCA	CPCA	PPCA	SVD_LPP	PBLDA	VPCA
1	0.81	0.76	0.83	0.85	0.64	0.60	**0.90**
2	0.64	**0.66**	**0.66**	0.60	0.55	0.51	**0.66**
3	**0.70**	0.68	0.68	0.62	0.53	0.53	0.68
4	**0.91**	0.89	0.85	0.84	0.86	0.78	**0.91**
5	0.62	0.63	**0.65**	0.60	0.58	0.51	0.63
6	**0.99**	0.48	**0.99**	0.50	**0.99**	0.82	0.78
7	0.54	0.46	0.50	**0.56**	0.53	0.52	0.50
8	0.30	0.30	0.29	0.30	**0.32**	0.29	0.28
9	0.78	0.73	0.78	0.73	0.57	0.21	**0.83**
10	0.98	0.67	0.78	0.67	0.87	**1.00**	0.96
11	0.79	0.55	0.76	0.56	0.57	0.60	**0.81**
12	**0.88**	0.44	0.59	0.47	0.78	0.42	0.59
13	**1.00**	0.69	0.70	0.69	0.64	0.51	0.73
14	0.44	0.35	0.41	0.34	**0.45**	0.33	0.36
15	0.91	0.58	0.95	0.55	0.60	0.11	**0.96**
16	**0.79**	0.75	**0.79**	0.50	0.75	0.50	—
17	**0.98**	**0.98**	**0.98**	0.97	**0.98**	0.93	—
18	0.96	0.89	**0.98**	0.73	0.75	0.93	—
19	**0.85**	0.69	0.80	0.62	0.62	0.65	—
20	**0.93**	0.31	0.49	0.57	0.73	0.50	—

对于其他几种方法，CPCA 方法相比于 PCA 方法在降维有效性上有较大提升。前者在 15 个数据集中的分类精度均比后者高，原因在于前者将 MTS 投影到公共低维子空间，后者则投影到不同低维子空间。但是，2 种方法仅对变量维度降维，序列长度没有缩减。PPCA 方法对参数 ω 比较敏感，当分段数量比较合理时，其降维有效性较好。原因在于分段数量将很大程度上决定最终计算得到的平均协方差矩阵，从而导致最终投影得到的降维序列差异较大。VPCA 方法分类精度较高，但只能用于等长数据集。PBLDA 方法从变量维度和时间维度进行降维，但在部分数据集上降维效果不佳。原因在于，面对不等长数据集时，PBLDA 方法采用截断方式将不等长序列转化为等长序列，造成信息损失。

SVD_LPP 方法从 MTS 中提取特征序列并进行 LPP 降维，解决了不等长问题，但只考虑了数据集局部信息，忽略了全局信息。并且，该方法需要对每个MTS进行奇异值分解，具有较高的计算复杂度。

另外，在多类别数据集 ASL 中，GLSUP 方法相比其他方法具有明显优势，其分类精度达到了 0.93，高于 SVD_LPP 分类精度的 0.73。原因在于 GLSUP 方法考虑了数据集全局和局部信息，在无标签数据集中也能将样本投影形成可分性较好的簇。

为了体现GLSUP方法降维效果，本节针对ASL数据集，可视化分析降维中的每一步骤。从数据集中选取 4 个样本，依次显示原始序列、特征序列以及降维序列，ASL 数据集降维效果的可视化结果如图 4 所示。括号内的数字为样本在数据集中的编号，英文为该样本所代表的语意（类别）。其中，GLSUP 方法降维之后的维度 d 取 10。

ASL 数据集中，1#序列和 2#序列属于同一类别，相似程度较高；1#序列和 55#序列属于不同类别，差异明显；1#序列和 28#序列属于不同类别，但是差异性较小。对于提取的等长特征序列，相同类别之间相似程度较高，而不同类别之间的差异性

较大。这说明本文提出的 MTS 特征提取方法能够刻画原始序列的有效特征，但提取后的特征序列长度较长，有 253 个维度。从图 4 可以看出，特征序列存在一定冗余信息，需要继续进行降维处理。经过 GLSUP 方法降维后，降维序列在仅有 10 个维度的情况下依然能够区分不同类别样本。

从数据集的降维可视化看出，GLSUP 方法具有较大的降维幅度，降维后的样本可分性较好。

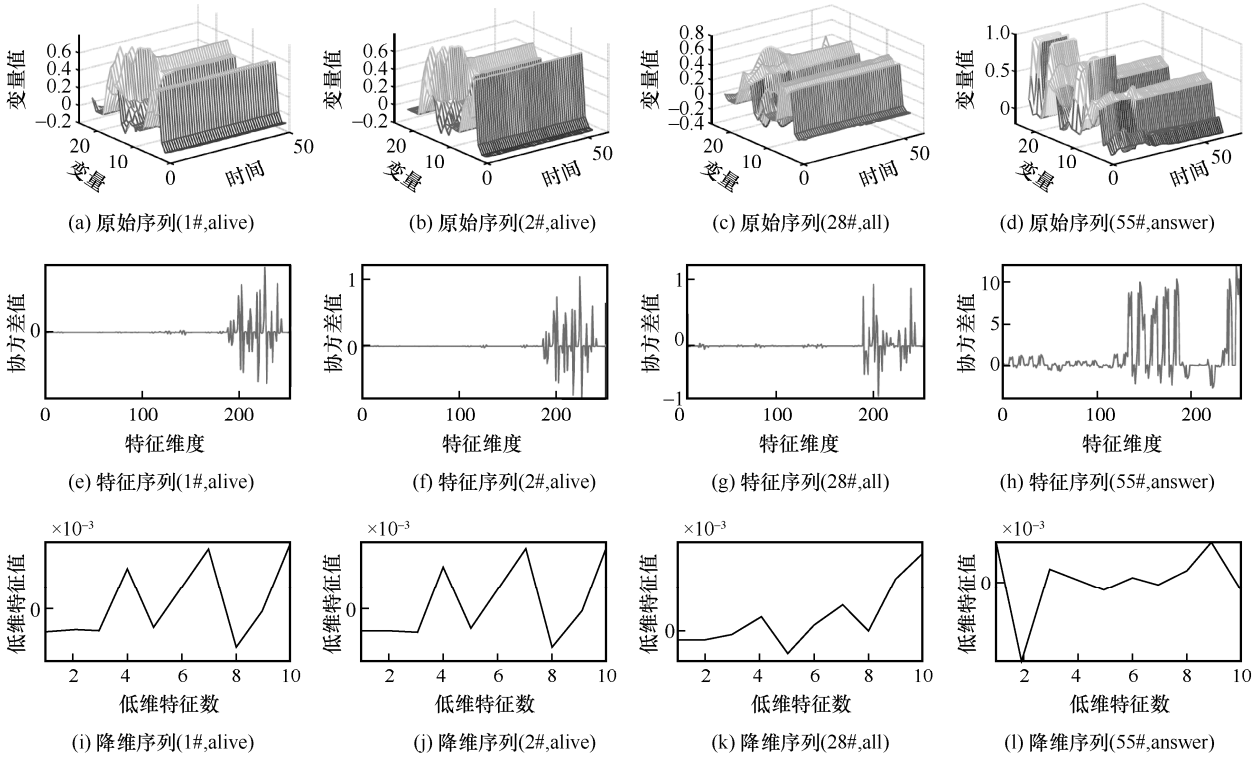
4.4　降维幅度比较

在上述降维有效性实验过程中，选取各种方法的最优参数，按照式(15)计算降维幅度，如表 3 所示。表 3 中每行的最大降维幅度用粗体表示，第二大降维幅度用下划线表示。SVD_LPP 方法在各数据集上的降维幅度均在 0.94 之上，原因在于该方法采用 SVD 方法从原始序列中提取得到特征序列，其维度等于变量数。GLSUP 也采用“两阶段”降维思想，通过提取的特征序列显著降低了原始 MTS 的维度，因此其降维幅度仅次于 SVD_LPP 方法，在 6 个数据集中取得了最大降维幅度，在 10 个数据集中取得了第二大降维幅度。同时，GLSUP 方法的降维有效性明显优于 SVD_LPP，如表 2 所示。原因在于 SVD_LPP 方法在构建图结构时只考虑了局部信息，忽略了全局信息。在变量数较少但序列长度较长的数据集上，GLSUP 方法降维幅度较大，这是因为提取得到的特征序列长度只与变量数有关，而与原始序列长度无关。

PCA、CPCA、PPCA、PBLDA、VPCA 方法均针对原始 MTS 进行降维，使降维幅度有限。此外，PCA、CPCA 和 PPCA 方法仅对变量维度进行降维，MTS 的序列长度没有缩短。VPCA 方法仅对 MTS 中的时间维度进行降维，变量数量没有减少。在 Epilepsy 数据集中，PCA 方法和 PPCA 方法的降维幅度为 0。原因为在方差贡献率为 80%的情况下，2 种方法降维后的变量数没有改变。在 RS、BM、Wafer 和 ASL 数据集中，PBLDA 方法降维幅度为 0，这是因为在这几个数据集中，PBLDA 特征维度 p_c 的最优参数值即数据集变量数。VPCA 仅能应用于等长 MTS 数据集，因此，在不等长数据集上没有给出降维幅度的实验结果。

4.5　参数敏感性分析

GLSUP 方法涉及的参数有近邻数 k 和降维之后的维度 d 。k 取值为 1、5、10、15、20，d 根据数据集的变量数间隔取值，参数敏感性分析结果如图 5 所示。在 DSA、HMD、Cricket、RS、Epilepsy、LSST、AWR 数据集上，GLSUP 方法对参数 d 比较敏感，这是因为参数 d 会直接影响降



(a) 原始序列(1#,alive)　(b) 原始序列(2#,alive)　(c) 原始序列(28#,all)　(d) 原始序列(55#,answer)

(e) 特征序列(1#,alive)　(f) 特征序列(2#,alive)　(g) 特征序列(28#,all)　(h) 特征序列(55#,answer)

(i) 降维序列(1#,alive)　(j) 降维序列(2#,alive)　(k) 降维序列(28#,all)　(l) 降维序列(55#,answer)

图 4　ASL 数据集降维效果的可视化结果

表 3　　降维幅度实验结果

数据集编号	GLSUP	PCA	CPCA	PPCA	SVD_LPP	PBLDA	VPCA
1	0.96	0.50	0.83	0.50	**0.97**	0.83	0.80
2	0.90	0.50	0.67	0.50	**0.94**	0.67	0.93
3	0.93	0.50	0.67	0.50	**0.94**	0.17	0.93
4	0.91	0.50	0.83	0.50	**0.94**	0.83	0.80
5	**0.94**	0.50	0.83	0.50	**0.94**	0.67	0.80
6	0.96	0.51	0.84	0.51	**1.00**	0.89	0.99
7	0.75	0.71	0.71	0.61	**1.00**	0.14	0.98
8	0.99	0.70	0.80	0.70	**1.00**	0.20	0.94
9	0.95	0.92	0.92	0.88	**0.99**	0.92	0.96
10	**1.00**	0.33	0.67	0.33	**1.00**	0.83	**1.00**
11	0.93	0.33	0.33	0.33	**0.97**	0.00	0.77
12	**1.00**	0.00	0.33	0.00	**1.00**	0.67	0.87
13	**1.00**	0.33	0.50	0.33	0.99	0.00	0.86
14	0.91	0.33	0.67	0.33	**0.97**	0.33	**0.97**
15	0.98	0.56	0.44	0.56	**0.99**	0.11	**0.99**
16	0.99	0.64	0.86	0.64	**1.00**	0.64	—
17	0.98	0.67	0.67	0.67	**0.99**	0.00	—
18	**1.00**	0.92	0.92	0.94	**1.00**	0.90	—
19	**1.00**	0.94	0.94	0.95	**1.00**	0.87	—
20	0.88	0.86	0.86	0.86	**0.98**	0.00	—

注：降维幅度 1.00 为保留两位小数的四舍五入结果

维幅度，从而对降维序列保留的信息产生影响，当 $d=1$ 时，分类精度较低，这是因为把维度较高的特征序列降低到 1 维时，丢失了较多信息。在 HMD、NATOPS、BM、EEGeye、WR、KP、ASL 数据集上，GLSUP 方法对参数 k 比较敏感，这是因为 k 值会影响 GLSUP 方法中的图结构，对降维结果有一定影响。

4.6　时间代价比较

时间代价分为两部分，即降维时间代价和分类时间代价。降维时间代价是指从原始 MTS 数据集得到降维特征集所耗费的时间。降维时间代价为获取特征序列、模型求解以及投影三部分的计算时间代价之和。选取以下 3 类典型数据集进行实验：样本规模较大的数据集 DSA、LSST，样本序列较长的数据集 HMD、Cricket，变量数较多的数据集 FM、NATOPS。各种方法的参数选取与 4.3 节相同。实验运行 10 次，时间代价取平均值，结果如表 4 所示。在表 4 中，对每一行最低时间代价用粗体表示，次低时间代价用下划线表示。

GLSUP 方法在 LSST、HMD、Cricket、NATOPS 等 4 个数据集中执行分类的时间代价为最低或次低，原因在于 GLSUP 方法降维幅度较大，显著降低了分类算法的计算代价。在样本规模较大、样本序列较长的数据集上，GLSUP 方法的总体时间代价较低，而在变量数较多的数据集上，时间代价优势不明显。下面从三类数据集的特点出发，分析具体原因。

在样本规模较大的数据集 DSA、LSST 上，GLSUP 方法的总时间代价较低。原因在于，该方法的计算复杂度与样本数量成线性关系，计算代价对样本数量有较好的适应性。

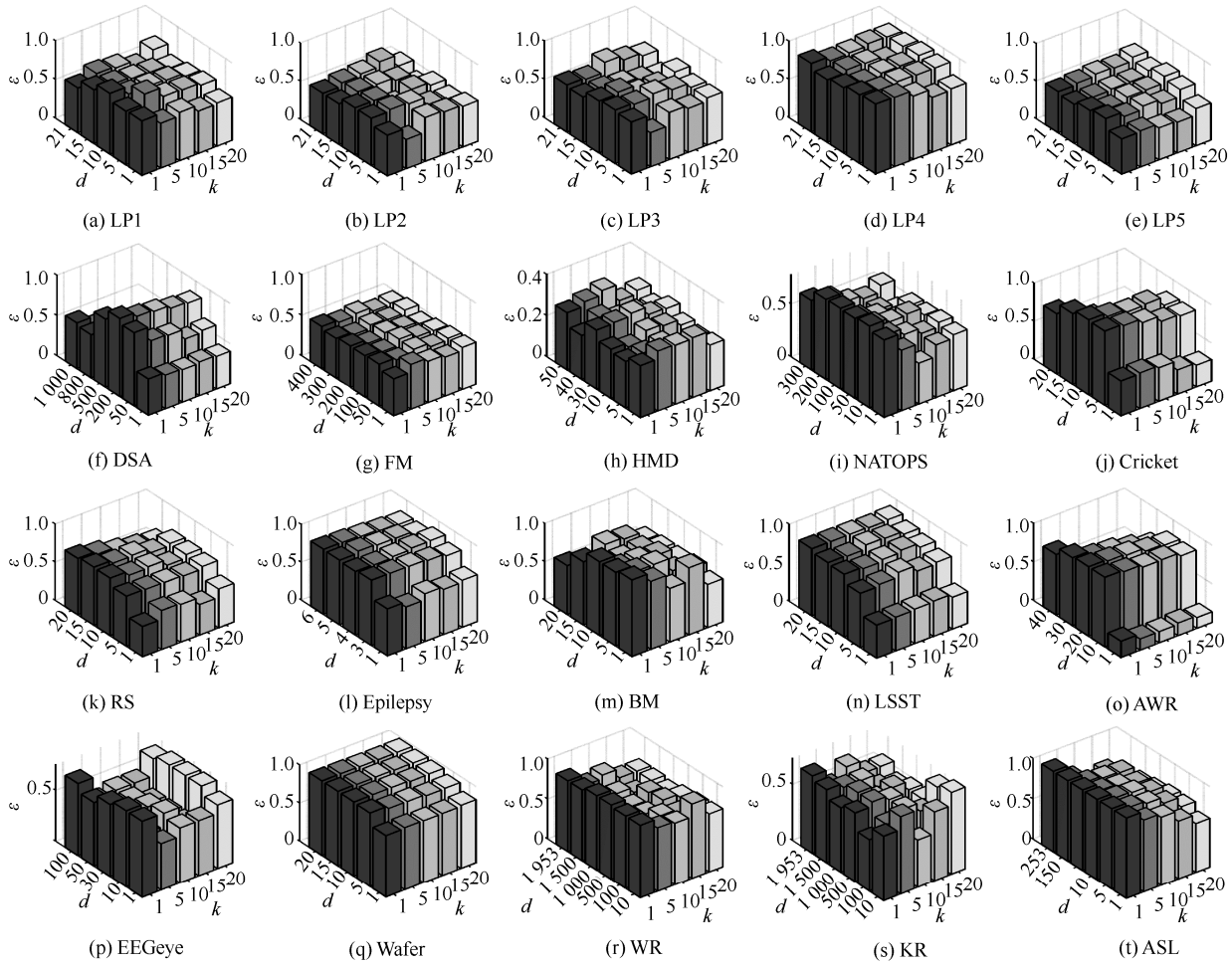


图 5　参数敏感性分析结果

表 4　　**时间代价实验结果**

数据集	GLSUP			PCA			CPCA			PPCA		
	降维时间代价/s	分类时间代价/s	总时间代价/s	降维时间代价/s	分类时间代价/s	总时间代价/s	降维时间代价/s	分类时间代价/s	总时间代价/s	降维时间代价/s	分类时间代价/s	总时间代价/s
DSA	17.47	2.68	20.15	1.35	386.24	387.59	**0.16**	82.85	83.01	417.77	414.56	832.32
LSST	0.25	7.55	**7.80**	0.09	108.52	108.61	**0.03**	54.41	54.43	0.15	130.87	131.03
HMD	0.04	**0.03**	**0.07**	0.02	1.93	1.95	**0.01**	1.51	1.52	0.02	1.42	1.44
Cricket	0.04	**0.01**	**0.05**	0.02	1.92	1.93	**0.01**	1.91	1.92	0.02	1.62	1.64
FM	0.77	0.21	0.98	0.05	1.83	1.87	**0.01**	1.81	1.82	0.06	2.73	2.79
NATOPS	0.24	**0.05**	0.29	0.03	0.33	0.36	**0.00**	0.33	0.34	0.04	0.58	0.62

数据集	SVD_LPP			PBLDA			VPCA		
	降维时间代价/s	分类时间代价/s	总时间代价/s	降维时间代价/s	分类时间代价/s	总时间代价/s	降维时间代价/s	分类时间代价/s	总时间代价/s
DSA	1.84	2.60	**4.44**	26.41	48.08	74.49	4.31	**1.96**	6.27
LSST	0.26	10.90	11.16	5.72	98.83	104.55	0.67	**7.19**	7.86
HMD	0.37	**0.03**	0.39	4.50	2.39	6.89	0.15	0.64	0.79
Cricket	1.23	0.02	1.25	5.75	0.09	5.84	0.73	0.10	0.82
FM	0.07	0.08	0.15	0.46	6.10	6.55	0.08	**0.06**	**0.14**
NATOPS	0.05	0.07	**0.11**	0.11	0.33	0.45	0.06	0.24	0.30

在样本序列较长的数据集 HMD、Cricket 上，GLSUP 方法的分类时间代价和总时间代价均最低。原因如下，从原始 MTS 中获取特征序列时，特征序列长度只与变量数有关，而与原始 MTS 的序列长度无关。降维计算复杂度与原始序列长度呈线性关系，因此对时间代价影响较小。

在变量数较多的数据集 FM、NATOPS 上，GLSUP 方法在降维环节的时间代价较高，其降维时间显著高于同属“两阶段”降维的 SVD_LPP 方法，如表 4 所示。原因如下，当 MTS 数据集变量数较多时，从原始 MTS 中提取的特征序列长度较长。导致局部散度矩阵 $\boldsymbol{S}_{\mathrm{L}}$ 和全局散度矩阵 $\boldsymbol{S}_{\mathrm{G}}$ 规模较大，在进行广义特征值求解时，时间代价较大。

5　结束语

目前，针对 MTS 的降维方法并不丰富，传统降维方法不能直接应用。现有的 MTS 降维方法难以在保证降维有效性的同时，较大幅度地降低数据维度。针对该问题，本文首先提出一种 MTS 特征序列提取方法，计算 MTS 的协方差矩阵，提取协方差矩阵的上三角元素，将其组合为特征序列，既保留了变量间的相关信息，又能够将数据集中所有等长或不等长 MTS 转化为等长的特征序列。然后，以“局部散度最小、全局散度最大”为基本思想，提出一种基于全局−局部散度的多元时间序列无监督降维方法，在保持局部近邻关系的同时，尽可能保留全局信息。实验结果表明，所提方法不仅具有较好的降维有效性，而且能够较大幅度地降低 MTS 维度。

本文提出的 MTS 降维方法以协方差矩阵为基础，体现变量间的线性关系，但不能刻画变量间的非线性关系。此外，所提方法虽对序列数量的变化具有较好的适应性，但当 MTS 变量维数较多时，提取的特征序列维度较高，增加了后续模型求解的计算代价。后续工作将针对以上问题开展持续研究。

参考文献：

[1] DHAR V, SUN C S, BATRA P. Transforming finance into vision: concurrent financial time series as convolutional nets[J]. Big Data, 2019, 7(4): 276-285.

[2] KANAVOS A, KOUNELIS F, ILIADIS L, et al. Deep learning models for forecasting aviation demand time series[J]. Neural Computing and Applications, 2021, 33(23): 16329-16343.

[3] 李正欣, 张凤鸣, 李克武, 等. 一种支持 DTW 距离的多元时间序列索引结构[J]. 软件学报, 2014, 25(3): 560-575.
LI Z X, ZHANG F M, LI K W, et al. Index structure for multivariate time series under DTW distance metric[J]. Journal of Software, 2014, 25(3): 560-575.

[4] MARIN Z P A, ROTH S, SCHMUTZLER D, et al. Self-supervised feature extraction from image time series in plant phenotyping using triplet networks[J]. Bioinformatics, 2021, 37(6): 861-867.

[5] ZHU H G, XIAO R Y, ZHANG J P, et al. A driving behavior risk classification framework via the unbalanced time series samples[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 2503312.

[6] CHEN Y, MANCHESTER W B, HERO A O, et al. Identifying solar flare precursors using time series of SDO/HMI images and SHARP parameters[J]. Space Weather, 2019, 17(10): 1404-1426.

[7] LIU Y, GAO J, CAO W, et al. A hybrid double-density dual-tree discrete wavelet transformation and marginal Fisher analysis for scoring sleep stages from unprocessed single-channel electroencephalogram[J]. Quantitative Imaging in Medicine and Surgery, 2020, 10(3): 766-778.

[8] 张伟, 王志海, 原继东, 等. 一种时间序列鉴别性特征字典构建算法[J]. 软件学报, 2020, 31(10): 3216-3237.
ZHANG W, WANG Z H, YUAN J D, et al. Time series discriminative feature dictionary construction algorithm[J]. Journal of Software, 2020, 31(10): 3216-3237.

[9] HUANG X, WU L, YE Y S. A review on dimensionality reduction techniques[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2019, 33(10): 1950017.

[10] RAY P, REDDY S S, BANERJEE T. Various dimension reduction techniques for high dimensional data analysis: a review[J]. Artificial Intelligence Review, 2021, 54(5): 3473-3515.

[11] JIA W K, SUN M L, LIAN J, et al. Feature dimensionality reduction: a review [J]. Complex & Intelligent Systems, 2022(8): 2663-2693.

[12] HE H, TAN Y H. Unsupervised classification of multivariate time series using VPCA and fuzzy clustering with spatial weighted matrix distance[J]. IEEE Transactions on Cybernetics, 2020, 50(3): 1096-1105.

[13] KUMAR G, SINGH U P, JAIN S. Hybrid evolutionary intelligent system and hybrid time series econometric model for stock price forecasting[J]. International Journal of Intelligent Systems, 2021, 36(9): 4902-4935.

[14] WAN X J, LI H L, ZHANG L P, et al. Dimensionality reduction for multivariate time-series data mining[J]. The Journal of Supercomputing, 2022, 78(7): 9862-9878.

[15] LIN W, HUANG J Z, MCELROY T. Time series seasonal adjustment using regularized singular value decomposition[J]. Journal of Business & Economic Statistics, 2020, 38(3): 487-501.

[16] LIU C H, JAJA J, PESSOA L. LEICA: Laplacian eigenmaps for group ICA decomposition of fMRI data[J]. NeuroImage, 2018, 169: 363-373.

[17] ZHAO J H, SUN F, LIANG H Y, et al. Pseudo bidirectional linear discriminant analysis for multivariate time series classification[J]. IEEE Access, 2021, 9: 88674-88684.

[18] POSPELOV N, TETEREVA A, MARTYNOVA O, et al. The Laplacian eigenmaps dimensionality reduction of fMRI data for discovering stimulus-induced changes in the resting-state brain activity[J]. Neuroimage: Reports, 2021, 1(3): 100035.

[19] WENG X Q, SHEN J Y. Classification of multivariate time series

using locality preserving projections[J]. Knowledge-Based Systems, 2008, 21(7): 581-587.

[20] WENG X Q. Classification of multivariate time series using supervised locality preserving projection[C]//Proceedings of the Third International Conference on Intelligent System Design and Engineering Applications. Piscataway: IEEE Press, 2013: 428-431.

[21] 董红玉, 陈晓云. 基于奇异值分解和判别局部保持投影的多变量时间序列分类[J]. 计算机应用, 2014, 34(1): 239-243.
DONG H Y, CHEN X Y. Classification of multivariate time series based on singular value decomposition and discriminant locality preserving projection[J]. Journal of Computer Applications, 2014, 34(1): 239-243.

[22] YAO B B, SU J, WU L F, et al. Modified local linear embedding algorithm for rolling element bearing fault diagnosis[J]. Applied Sciences, 2017, 7(11): 1178.

[23] 胡钢, 李正欣, 张凤鸣, 等. 二维类间边界 Fisher 分析的多元时间序列降维[J]. 北京航空航天大学学报, 2023, 49(12): 3537-3546.
HU G, LI Z X, ZHANG F M, et al. Dimension reduction of multivariate time series based on two-dimensional inter-class marginal Fisher analysis[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(12): 3537-3546.

[24] KARAMITOPOULOS L, EVANGELIDIS G, DERVOS D. PCA-based time series similarity search[M]. Berlin: Springer, 2010.

[25] 李正欣, 郭建胜, 惠晓滨, 等. 基于共同主成分的多元时间序列降维方法[J]. 控制与决策, 2013, 28(4): 531-536.
LI Z X, GUO J S, HUI X B, et al. Dimension reduction method for multivariate time series based on common principal component[J]. Control and Decision, 2013, 28(4): 531-536.

[26] LI H L. Accurate and efficient classification based on common principal components analysis for multivariate time series[J]. Neurocomputing, 2016, 171: 744-753.

[27] 李正欣, 张凤鸣, 张晓丰, 等. 多元时间序列特征降维方法研究[J]. 小型微型计算机系统, 2013, 34(2): 338-344.
LI Z X, ZHANG F M, ZHANG X F, et al. Research on feature dimension reduction method for multivariate time series[J]. Journal of Chinese Computer Systems, 2013, 34(2): 338-344.

[28] SUNDARARAJAN R R. Principal component analysis using frequency components of multivariate time series[J]. Computational Statistics & Data Analysis, 2021, 157: 107164.

[29] 李海林. 基于变量相关性的多元时间序列特征表示[J]. 控制与决策, 2015, 30(3): 441-447.
LI H L. Feature representation of multivariate time series based on correlation among variables[J]. Control and Decision, 2015, 30(3): 441-447.

[30] NIE F P, ZHU W, LI X L. Unsupervised large graph embedding based on balanced and hierarchical K-means[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(4): 2008-2019.

[31] WANG Z, ZHANG L, WANG B J. Sparse modified marginal fisher analysis for facial expression recognition[J]. Applied Intelligence, 2019, 49(7): 2659-2671.

[32] ZHANG S J, MA Z M, ZHANG G K, et al. Dimensionality reduction based on multilocal linear pattern preservation[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(4): 1696-1709.

[33] WANG H, NIE F P, HUANG H. Globally and locally consistent unsupervised projection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2014, 28(1): 1328-1333.

[34] CHEN J H, WAN Y, WANG X Y, et al. Learning-based shapelets discovery by feature selection for time series classification[J]. Applied Intelligence, 2022, 52(8): 9460-9475.

[35] BAGNALL A, DAU H A, LINES J, et al. The UEA multivariate time series classification archive[J]. arXiv Preprint, arXiv: 1811.00075, 2018.

[36] MARKELLE K, RACHEL L, KOLBY N. The UCI machine learning repository[R]. 2023.

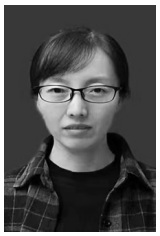
[作者简介]

李正欣（1982− ），男，河南信阳人，博士，空军工程大学副教授、硕士生导师，主要研究方向为时间序列模式识别、数据挖掘与机器学习等。

胡钢（1998− ），男，江西九江人，空军工程大学博士生，主要研究方向为时间序列降维、数据挖掘与机器学习等。

张凤鸣（1963− ），男，重庆人，空军工程大学教授、博士生导师，主要研究方向为信息系统工程与智能决策。

张晓丰（1978− ），男，天津人，博士，空军工程大学副教授、硕士生导师，主要研究方向为信息系统工程与智能决策。



赵永梅（1982− ），女，陕西延安人，博士，空军工程大学副教授，主要研究方向为数据融合与补全、信息物理系统等。