

# 基于长期注意力的多维时间序列形态表征<sup>①</sup>

朱潇凌<sup>1,2</sup>, 陈黎飞<sup>1,2,3</sup>

<sup>1</sup>(福建师范大学 计算机与网络空间安全学院, 福州 350117)

<sup>2</sup>(福建师范大学 数字福建环境监测物联网实验室, 福州 350117)

<sup>3</sup>(福建师范大学 统计学与人工智能福建省高校重点实验室, 福州 350117)

通信作者: 陈黎飞, E-mail: [clfei@fjnu.edu.cn](mailto:clfei@fjnu.edu.cn)



**摘要:** 多维时间序列数据广泛应用于多个领域, 对其有效的表征对下游分析与挖掘任务至关重要. 传统形态元变换方法通过将单维时间序列投影到形态元空间再行融合的方式进行特征提取, 未考虑不同维度时序间的复杂的耦合关系, 同时形态元长度的限制也妨碍了序列上长程依赖关系的获取. 针对这些问题, 提出了一种耦合维度依赖与长程依赖的多维时序表征方法 (CDT-ShapeNet). 该方法中维度信息表征模块通过维度注意力机制捕捉不同维度之间的依赖关系, 而长期信息表征模块则利用注意力机制和长短期记忆网络学习长期时间依赖. 在 9 个 UEA 数据集上进行实验, 结果显示, 较之于对比方法, 平均准确率提高了 6.8%, 验证了其在多维时间序列表征方面的有效性.

**关键词:** 多维时间序列; 维度依赖; 长程依赖; 时间序列表征

引用格式: 朱潇凌, 陈黎飞. 基于长期注意力的多维时间序列形态表征. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9860.html>

## Shapelet Representation for Multidimensional Time Series Based on Long-term Attention

ZHU Xiao-Ling<sup>1,2</sup>, CHEN Li-Fei<sup>1,2,3</sup>

<sup>1</sup>(College of Computer and Cyber Security, Fujian Normal University, Fuzhou 350117, China)

<sup>2</sup>(Digit Fujian Internet-of-Things Laboratory of Environmental Monitoring, Fujian Normal University, Fuzhou 350117, China)

<sup>3</sup>(Fujian Provincial Key Laboratory of Statistics and Artificial Intelligence, Fujian Normal University, Fuzhou 350117, China)

**Abstract:** Multidimensional time series data are widely used across various fields, and their effective representation is critical for subsequent analysis and mining tasks. Traditional shapelet transform methods extract features by projecting the single-dimensional time series into the shapelet space and then fusing them without considering the complex coupling relationships between different dimensions. Moreover, the restriction on shapelet length hinders the acquisition of long-range dependencies on sequences. To address these issues, a multidimensional time series representation method, CDT-ShapeNet, coupling both dimensional dependencies and long-range dependencies is proposed in this study. In this method, the dimensional information representation module captures the dependencies between different dimensions through a dimensional attention mechanism, while the long-term information representation module learns long-term temporal dependencies using an attention mechanism and a long-short-term memory network. Experiments conducted on nine UEA datasets show that this method enhances the average accuracy by 6.8% in comparison with other methods, validating its effectiveness in multidimensional time series representation.

**Key words:** multidimensional time series; dimensional dependence; long-range dependence; time series representation

<sup>①</sup> 基金项目: 福建省自然科学基金 (2024J01067); 国家自然科学基金 (U1805263)

收稿时间: 2024-10-24; 修改时间: 2024-11-18, 2024-12-09; 采用时间: 2024-12-18; csa 在线出版时间: 2025-03-07

多维时间序列指由多个变量按时间顺序排列而成的数据集合,即在同一时间点上存在多个观测值。多维时间序列数据广泛应用于诸多领域,如语音识别、多媒体处理、医学诊断、经济分析、科学研究和工程实践等,每天都会生成大量此类数据<sup>[1,2]</sup>。时间序列表征通过某种方法,找到时间序列潜藏的结构,生成特征向量,并为后续工作提供支持<sup>[3,4]</sup>。多维时间序列分类是多维时间序列研究中的一个关键问题,其目的是通过分析已知时间序列数据与标签之间的关系,挖掘潜在模式,从而对未知数据进行有效分类<sup>[5]</sup>。多维时间序列分类中的一个核心是找到有效的时间序列表征。由于多维时间序列包含多个维度,数据所蕴含的信息更加丰富,各维度间以及不同时间点之间存在复杂的相互关系,因此,发现多维时间序列分类中的表征充满挑战<sup>[6]</sup>。

多维时间序列分类中的表征方法可以分为两类,隐式表征方法和显式表征方法<sup>[7]</sup>。隐式表征方法通常基于相似性度量或网络自动学习来捕捉时间序列特征。相似性度量方法通过某种距离度量函数来衡量时间序列的之间相似性,进而对序列进行区分。欧几里得距离的方法<sup>[8]</sup>,通过欧几里得距离度量不同序列相似性。动态时间规整(dynamic time warping, DTW)<sup>[9-11]</sup>通过点对点的强制对齐,让不同长度时间序列可比。相似性度量方法由于距离公式的限制没有考虑维度之间关系。网络学习的方法,如多通道深度卷积神经网络(multi-channel deep convolutional neural network, MCDCNN)<sup>[12]</sup>通过拼接卷积网络获得每个维度的时间序列潜在特征。Karim 等人<sup>[13]</sup>提出了一个将长短期记忆网络(long short-term memory, LSTM)、卷积网络和挤压激励模块结合的方法来获取潜在特征。TapNet 方法<sup>[14]</sup>提出了一个注意力原型网络来学习低维特征并解决训练标签有限问题。但是深度学习方法获取的特征可解释性差。

显式表征方法通常将多维时间序列映射到新的特征空间,以提取更具代表性的变换域特征。如基于字典的方法通过滑动窗口从时间序列中提取子列,并将其转化为符号化的“单词”,利用这些符号进行特征挖掘。其中,时间序列单词提取(word extraction for time series classification, WEASEL)与多元符号扩展(multivariate unsupervised symbols, MUSE)的结合方法<sup>[15]</sup>,通过特定的特征推导技术来提取和过滤多元特征。然而,这类方法由于仅统计频率等信息,缺乏对时间序列长程依赖关系的捕捉。基于形态元(shapelets)的方法最

早由 Ye 等人提出<sup>[16]</sup>,该方法将时间序列与形态元的距离作为特征,并利用决策树进行分类。形态元直接来自原始序列,具有良好的可解释性。尤其是自形态元变换(shapelet transform)方法提出后<sup>[17]</sup>,该类方法不仅能够与决策树结合,还可以与支持向量机(SVM)和神经网络等模型结合,从而推动了这类方法的快速发展。形态元变换最初用于一维时间序列,随后也有一些变种方法将其扩展至多维时间序列中<sup>[18,19]</sup>,这些方法通常仅从某些维度中提取形态元<sup>[20,21]</sup>,没有充分考虑维度之间的耦合关系。此外,还有一些与深度学习相结合的形态元方法,如 ShapeNet<sup>[22]</sup>,这些方法虽然在特定任务中取得了一定进展,但仍然存在两个主要问题。首先,由于距离计算公式的限制,未能有效捕捉多维时间序列中不同维度之间的复杂的耦合关系。其次,形态元只是整个时间序列中的一个子列,属于局部特征,导致其提取的特征缺乏对时间序列长期信息的捕捉与建模。

针对上述挑战,本文提出一种名为 CDT-ShapeNet 的多维时间形态表征方法。首先,本文设计了一个维度信息表征模块,该模块通过维度注意力机制,有效地捕捉多维时间序列中各维度之间的相互依赖关系。在此基础上,该模块获得形态元投影空间维度依赖性的耦合关系。其次,引入了长期信息表征模块,该模块融合了时序注意力机制与 LSTM,以学习时间序列中的长程依赖关系。具体而言,时序注意力机制用于捕捉时间上的长期依赖性,而 LSTM 则通过其门控机制筛选出关键的时间信息,进一步提升模型的表征能力。最终,本文设计的特征融合层通过卷积和注意力机制整合多种特征。通过基于 UEA 数据集的分类实验作为下游任务,CDT-ShapeNet 模型在分类任务中表现出优越的性能,验证了其在多维时间序列中表征能力。

## 1 相关工作

在时间序列表征方法中,形态元变换方法将时间序列投影到形态元空间以提取时间序列表征。同时,LSTM 利用多个门结构获取长短期时间信息。

### 1.1 相关定义

在本文中,多维时间序列数据集被定义为  $T = \{(T_i, c_i), i = 1, 2, \dots, n\}$ , 其中  $n$  为样本个数,  $c_i \in C$  为样本类标签,  $C$  为有限的类别标签集合。  $T_i = \{t_1, t_2, \dots, t_L\} \in R^{m \times L}$  为一个  $m$  维  $L$  长的时间序列样本。  $t_p = \{t_{1,p}, t_{2,p}, \dots, t_{m,p}\}$ ,  $p \in \{1, 2, \dots, L\}$  为在时间点  $p$  的一个  $m$  维向量。将一个

形态元序列定义为  $S_j = \{s_1, s_2, \dots, s_W\} \in \mathbb{R}^{m \times W}$ ,  $S_j$  是一个  $m$  维  $W$  长的序列,  $s_p = \{s_{1,p}, s_{2,p}, \dots, s_{m,p}\}$ ,  $p \in \{1, 2, \dots, W\}$  为  $S_j$  时间点  $p$  上的一个  $m$  维向量.

## 1.2 形态元变换

形态元变换方法首先通过选取  $K$  个形态元 ( $K$  为预定义参数), 并根据特定的距离公式将原始时间序列投影到形态元空间中, 从而在该空间上获得新的特征表示. 该方法最初应用于一维时间序列, 随后扩展至多维时间序列的分析中. 在多维时间序列处理中, 依据形态元的维度选择和投影距离的计算方式, 可将形态元变换方法划分为 3 类: 独立形态元变换、多维相关形态元变换和多维无关形态元变换<sup>[23]</sup>. 独立形态元变换方法在多维时间序列中只选取一个维度, 并按照一维时间序列的方式对该维度的特征进行计算, 进而作为多维时间序列的特征. 多维相关形态元变换方法选取所有维度, 并保证各个维度之间相位一致. 在计算距离时, 需对每个维度同时滑动, 并将所有维度上的距离进行求和, 具体距离如式 (1) 所示:

$$dist_I = \min_{\forall p \in \{1, 2, \dots, L-W+1\}} \sum_{k=1}^m \sqrt{\sum_{i=1}^W \|s_{k,i} - t_{k,p+i}\|^2} \quad (1)$$

多维无关形态元变换方法也提取所有维度, 但各个维度之间没有依赖关系, 不要求相位一致. 计算距离时, 各个维度独立计算距离, 并取最小值作为距离, 距离计算如式 (2) 所示:

$$dist_D = \min_{\forall k \in \{1, 2, \dots, m\}} \min_{\forall p \in \{1, 2, \dots, L-W+1\}} \sqrt{\sum_{i=1}^W \|s_{k,i} - t_{k,p+i}\|^2} \quad (2)$$

由上述描述可知, 独立形态元变换方法假设仅某一个维度起作用, 而多维相关形态元变换方法则假设所有维度的作用是等同的. 相比之下, 多维无关形态元变换方法认为各维度之间相互独立. 然而, 这些方法在处理维度间关系时过于简化, 未能充分挖掘不同维度之间潜在的耦合关系.

## 1.3 LSTM

LSTM 利用输入门、遗忘门和输出门这 3 种门结构构造长短期记忆单元, 根据输入和门结构决定早期时间信息是否使用. LSTM 工作原理如下<sup>[24,25]</sup>.

### (1) 遗忘门

$$f_t = \sigma(U_f \times X_t + W_f \times h_{t-1} + b_f) \quad (3)$$

其中,  $U_f$  和  $W_f$  为权重参数,  $b_f$  为偏置项.  $X_t$  为当前时刻输入,  $h_{t-1}$  表示上一时刻短期信息状态.  $\sigma$  代表 Sigmoid 函数.

### (2) 输入门

$$i_t = \sigma(U_i \times X_t + W_i \times h_{t-1} + b_i) \quad (4)$$

$$g_t = \tanh(U_g \times X_t + W_g \times h_{t-1} + b_g) \quad (5)$$

$$C_t = C_{t-1} \odot f_t + i_t \odot g_t \quad (6)$$

其中,  $U_i$ 、 $W_i$ 、 $U_g$  和  $W_g$  为权重参数,  $b_i$  和  $b_g$  为偏置项,  $\odot$  为 Hadarmad 乘积.  $C_{t-1}$  表示上一时刻长期信息状态.

### (3) 输出门

$$O_t = \sigma(U_o \times X_t + W_o \times h_{t-1} + b_o) \quad (7)$$

$$h_t = O_t \odot \tanh(C_t) \quad (8)$$

其中,  $U_o$  和  $W_o$  为权重参数,  $b_o$  为偏置项,  $C_t$  为下一时刻长期信息状态.

LSTM 模型具备信息选择和提取时间依赖关系的能力, 它能够记忆旧有数据, 避免了因新信息输入使得历史上有效信息丢失. 该模型能够使时间序列中的关键信息得到有效更新和传递, 从而解决时间序列的长期依赖问题.

## 2 耦合维度依赖和长程依赖的表征方法

### 2.1 方法概述

形态元变换方法提取的特征没有考虑耦合维度之间依赖关系和时间之间的依赖关系, 但本方法分别提取耦合了维度依赖关系的特征和耦合了长程依赖关系的特征, 并将这两个特征结合得到耦合了维度依赖关系和长程依赖关系的特征. 本方法的网络结构如图 1 所示, 主要由 3 部分组成. 第 1 部分, 维度信息表征模块学习形态元变换之间复杂维度关系获得新的形态元变换特征. 第 2 部分, 长期信息表征模块学习时间序列在形态元空间上长期信息. 第 3 部分, 通过特征融合层将两者结合, 使得新的特征既有维度之间信息又有长期信息. 最后通过输出层得到下游任务输出结果.

### 2.2 维度信息表征模块

多维时间序列的各维度之间存在依赖关系, 维度信息表征模块旨在捕获时间序列投影到形态元空间后的维度依赖特征. 该模块由两个部分组成: 其一是形态

元投影层,用于将时间序列投影到形态元空间中,从而获得其投影特征;另一个是维度依赖层,进一步处理这些投影特征,提取并增强其中包含的维度依赖关系特征.这一设计捕捉了多维时间序列中不同维度间的依赖性.

$$ST_{i,b} = \min_{p \in \{1,2,\dots,L-W+1\}} \sum_{j=1}^W \sqrt{\|s_{b,j} - t_{b,p+j}\|^2} \quad (9)$$

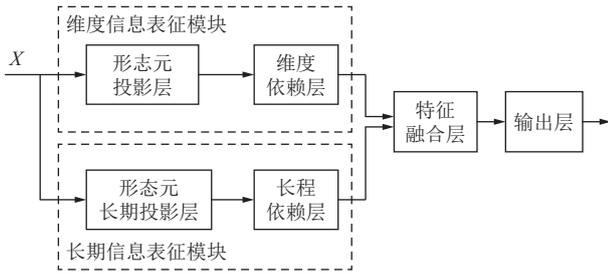


图1 CDT-ShapeNet 网络结构

如式(9)所示通过计算某个维度在  $S_j$  同一维度下的最小值获得  $T_i$  的第  $b$  维在  $S_j$  上相应投影  $ST_{i,b}$ .

$$ST_i = (ST_{i,1}, ST_{i,2}, \dots, ST_{i,m}) \quad (10)$$

而  $T_i$  在  $S_j$  上的投影如式(10)所示.通过形态元投影层得到的投影特征未考虑耦合维度间的依赖关系,因此可以采用注意力机制来挖掘这些维度之间的依赖关系.首先,将投影特征映射到新的子空间,得到两个与维度相关的向量  $Q_{\text{dim}}$  和  $K_{\text{dim}}$ :

$$Q_{\text{dim}} = W_{\text{dim},Q} \times ST_i \quad (11)$$

$$K_{\text{dim}} = W_{\text{dim},K} \times ST_i \quad (12)$$

$$V_{\text{dim}} = W_{\text{dim},V} \times ST_i \quad (13)$$

投影特征矩阵通过线性变换得到  $Q_{\text{dim}}$  和  $K_{\text{dim}}$  两个相关性向量以及值矩阵  $V_{\text{dim}}$ ,其中,  $W_{\text{dim},Q}$ 、 $W_{\text{dim},K}$  和  $W_{\text{dim},V}$  是可学习的权重矩阵.接下来利用加性注意力计算方法来计算维度之间相关性的得分矩阵  $A_{\text{dim}}$ :

$$A_{\text{dim}} = \text{Softmax}(W_{\text{dim},s} \times \tanh(Q_{\text{dim}} + K_{\text{dim}})) \quad (14)$$

$$ST_{i,\text{dim}} = A_{\text{dim}} \times V_{\text{dim}} \quad (15)$$

相关性得分矩阵  $A_{\text{dim}}$  由维度相关性向量  $Q_{\text{dim}}$  和  $K_{\text{dim}}$  计算得到,其中  $W_{\text{dim},s}$  为权重参数.这里  $A_{\text{dim}}$  用于量化不同维度之间的依赖关系.通过  $A_{\text{dim}}$  加权  $V_{\text{dim}}$  得到融合了维度依赖关系的新投影空间特征  $ST_{i,\text{dim}}$ .

综上,通过维度信息表征模块,获得了多维时间序

列维度之间的依赖关系,并得到耦合了维度依赖关系的新特征.

### 2.3 长期信息表征模块

多维时间序列除了维度之间的依赖关系还有时间上的长程依赖关系.因此,提出长期信息表征模块捕获时间上的长程依赖关系.形态元长期投影层首先根据式(16)计算出时间序列  $T_i$  第  $b$  维第  $p$  个起始点在形态元  $S_j$  上的投影  $GT_{i,b,p}$ :

$$GT_{i,b,p} = \sum_{j=1}^W \sqrt{\|s_{b,j} - t_{b,p+j}\|^2} \quad (16)$$

每个  $GT_{i,b,p}$  代表了时间序列上  $p$  点  $W$  长范围内的信息,将  $S_j$  在时间序列上滑动就能获得所有点信息,因而第  $b$  维上整个时间的信息  $GT_{i,b}$  如式(17)所示:

$$GT_{i,b} = (GT_{i,b,1}, GT_{i,b,2}, \dots, GT_{i,b,L-W+1}) \quad (17)$$

之后获得所有维度上整个时间的时序矩阵  $GT_i$ .此时获得的特征包含了粗略的长期时间信息,没有挖掘长期时间中的长程依赖关系.

$$GT_i = (GT_{i,1}, GT_{i,2}, \dots, GT_{i,m}) \quad (18)$$

通过长程依赖层对此特征进行挖掘提取,得到融入了长期时间依赖关系的特征.长程依赖层利用时序注意力机制获得不同时间点之间的依赖关系,首先,通过线性变换将时间序列矩阵  $GT_i$  转换,得到相关性向量  $Q_{\text{long}}$  和  $K_{\text{long}}$  以及值矩阵  $V_{\text{long}}$ :

$$Q_{\text{long}} = W_{\text{long},Q} \times GT_i \quad (19)$$

$$K_{\text{long}} = W_{\text{long},K} \times GT_i \quad (20)$$

$$V_{\text{long}} = W_{\text{long},V} \times GT_i \quad (21)$$

其中,  $W_{\text{long},Q}$ 、 $W_{\text{long},K}$  和  $W_{\text{long},V}$  为权重参数.时序注意力机制采用点积注意力计算方法得到不同时间点之间的相似性权重  $A_{\text{long}}$ ,如式(22)所示:

$$A_{\text{long}} = \text{Softmax}(Q_{\text{long}} \times K_{\text{long}}^T) \quad (22)$$

其中,  $A_{\text{long}}$  用于量化不同时间点之间的相似性.通过  $A_{\text{long}}$  加权  $V_{\text{long}}$  进而得到长程依赖的特征  $value$ ,如下所示:

$$value = A_{\text{long}} \times V_{\text{long}} \quad (23)$$

利用注意力机制发掘了长程依赖关系,其中包含丰富时间信息,而 LSTM 使用门结构控制上一时刻信息使用情况,使得过去信息不断融入当前信息中,最后

得到包含了长程依赖的特征. 因而利用 LSTM 得到了融合了长程依赖关系的特征  $ST_{i,\text{long}}$ .

## 2.4 特征融合层

上述两个模块都是从形态元空间中提取信息, 因而存在一定的相关性. 然而, 维度信息表征模块侧重于挖掘维度间的关系信息, 而长期信息表征模块则关注时间序列中的长程依赖关系. 特征融合层通过计算特征之间的相似性, 使相似的部分更好地融合, 从而获得更好的特征表示.

$$F_{\text{dim}} = W_{\text{dim},F} \times ST_{\text{dim}} \quad (24)$$

$$F_{\text{long}} = W_{\text{long},F} \times ST_{\text{long}} \quad (25)$$

首先通过两个核大小为 1 的卷积操作, 将之前获得的耦合维度依赖的特征和长程依赖的特征映射为两个新特征, 如式 (24) 和式 (25) 所示, 其中  $W_{\text{dim},F}$  和  $W_{\text{long},F}$  为卷积核参数.

$$F_f = \text{Softmax}(F_{\text{dim}} \times F_{\text{long}}^T) \times (W_f \times ST_{\text{long}}) \quad (26)$$

然后如式 (26) 所示, 通过矩阵乘法计算两个特征之间相似性, 并通过 *Softmax* 函数得到权重参数, 并以此为依据对特征进行调整得到融合后特征, 其中  $W_f$  为权重参数. 再通过连接方式将  $F_f$  和  $ST_{\text{dim}}$  融合, 如式 (27) 所示:

$$O_f = \text{Concat}(F_f, ST_{\text{dim}}) \quad (27)$$

最后, 将融合后的特征通过一个由全连接层构成的输出层得到输出结果  $f(T_i)$ .

## 2.5 损失函数

考虑到下游任务是分类任务, 采用负对数似然损失衡量分类损失. 为了使学习到的特征更加集中, 采用中心损失. 总损失通过两者相结合的方式得到. 负对数似然损失如式 (28) 所示:

$$L_{\text{nil}}(p(y|x), y) = -\log f(T_i) \quad (28)$$

中心损失如式 (29) 所示, 其中  $c$  为中心点特征,  $x_i$  为样本特征.

$$L_{\text{center}} = \frac{1}{2n} \sum_{i=1}^n \|x_i - c\|^2 \quad (29)$$

其中, 针对耦合了维度依赖的特征  $ST_{i,\text{dim}}$  的中心损失如式 (30) 所示:

$$L_{\text{center},\text{dim}} = \frac{1}{2n} \sum_{i=1}^n \|ST_{i,\text{dim}} - c_{\text{dim}}\|^2 \quad (30)$$

针对耦合了长程依赖的特征  $ST_{i,\text{long}}$  的中心损失如式 (31) 所示.

$$L_{\text{center},\text{long}} = \frac{1}{2n} \sum_{i=1}^n \|ST_{i,\text{long}} - c_{\text{long}}\|^2 \quad (31)$$

最后总损失为  $L$  如式 (32) 所示, 其中  $L_{\text{center},\text{dim}}$  为维度信息表征模块学习到特征中心损失,  $L_{\text{center},\text{long}}$  为长期信息表征模块学习到特征中心损失.

$$L = L_{\text{nil}} + \lambda_1 L_{\text{center},\text{dim}} + \lambda_2 L_{\text{center},\text{long}} \quad (32)$$

使用反向传播来优化  $f(T_i)$ 、 $ST_{i,\text{dim}}$ 、 $c_{\text{dim}}$ 、 $ST_{i,\text{long}}$  和  $c_{\text{long}}$ , 并更新网络参数. 在本文中两个参数  $\lambda_1$  和  $\lambda_2$ , 通过使用交叉验证实验得出, 最终选用都为 0.1.

## 3 实验分析

本节主要验证提出的耦合维度依赖和长程依赖的表征方法, 在 UEA 数据集<sup>[26]</sup>上与若干多维时间序列表征算法进行了对比.

### 3.1 实验数据集

本研究采用了 UEA 多维时间序列分类基准数据集作为实验对象. UEA 数据集广泛应用于时间序列分类问题研究, 涵盖多个领域, 包括人类活动识别、心电图分析、音频频谱处理等, 为多维时间序列模型的评估提供了丰富且具有挑战性的测试环境. 本次实验共选择 9 个数据集, 每个数据集的特点和属性如表 1.

表 1 数据集信息

数据集名称	领域	样本数量	长度	维数
AtrialFibrillation	心电图	30	640	2
BasicMotions	人类活动识别	80	100	6
Cricket	人类活动识别	180	1197	6
FingerMovements	人类活动识别	416	50	28
HandMovementDirection	脑电图	234	400	10
Heartbeat	音频频谱	409	405	61
SelfRegulationSCP2	脑电图	380	1152	7
StandWalkJump	心电图	27	2500	4
UWaveGestureLibrary	人类活动识别	440	315	3

**AtrialFibrillation:** 包含双通道心电图 (ECG) 记录, 旨在预测房颤事件.

**BasicMotions:** 记录学生进行不同活动 (如步行、跑步等) 时的运动数据, 包含 4 个类别.

**Cricket:** 记录板球比赛裁判通过手势传达比赛信息的信号, 手势通过手部动作传递.

**FingerMovements**: 记录受试者使用食指和小指按下对应按键的动作序列, 任务为预测动作顺序.

**HandMovementDirection**: 通过脑磁图 (MEG) 记录腕部在 4 个不同方向上的运动数据.

**Heartbeat**: 从人体不同部位记录心脏相关声音, 分为正常和异常两类.

**SelfRegulationSCP2**: 记录接受人工呼吸的患者进行皮层电位调节的任务.

**StandWalkJump**: 记录 25 岁男性在站立、行走和跳跃等活动时的短时心电图信号.

**UWaveGestureLibrary**: 记录通过加速度计生成的 8 种简单手势数据, 用于手势分类任务.

### 3.2 对比算法

多维时间序列表征方法可以分为两大类, 分别为显示表征方法和隐式表征方法. 因此, 在这两类方法中分别选取代表性方法.

主流显示表征方法可以分为两类, 第 1 类是基于字典的方法, 本实验选取了 WEASEL+MUSE<sup>[15]</sup>, 该方法通过融合统计特征选择、可变窗口长度和符号化时间序列技术得到时间序列特征. 第 2 类显示表征方法选取 ShapeNet<sup>[22]</sup>方法, 该方法基于形态元变换, 通过编码器将不同维度和不同长度的形态元变换特征统一到同一空间, 使得变换后特征可比.

主流的隐式方法也分为两种类型. 其中基于距离度量方法中, 选取了 DTWI 和 DTWD<sup>[9]</sup>, 这两个算法都是基于 DTW 距离的, 但是距离函数是不同的, DTWI 距离计算时认为所有维度是独立的, 但 DTWD 距离计算时则认为维度之间相关. 另一类方法是基于网络学习的方法, 选取了 3 种采用不同神经网络中代表性的模块方法. MLSTM-FCNs<sup>[13]</sup>方法将 LSTM、卷积网络和挤压激励模块相结合提取时间序列特征. TapNet<sup>[14]</sup>

采用注意力原型模块将深度学习和传统方法各自优势相结合提取特征. 最后是新近提出的一个方法 MF\_NET<sup>[27]</sup>, 该方法将注意力机制与图神经网络相结合, 用于挖掘局部-全局-空间的时间序列特征.

### 3.3 实验设置

为了评估时间序列表征优异性, 将各个方法应用于分类任务, 并采用分类准确率作为评价指标. 数据集在实验前经过归一化处理, 训练集与测试集的划分与原始数据保持一致. 实验的最大迭代次数设定为 700 轮, 使用初始学习率为 1E-3 的 Adam 优化器进行优化. 所有训练过程均在 NVIDIA GeForce RTX 4090 GPU 上进行. 此外, 超参数 (如形态元个数、形态元长度和中心损失函数的权重) 通过实验调优确定, 以确保模型的最佳性能.

### 3.4 实验结果

为了验证本文方法, 表 2 展示了本文方法与其他方法在 UEA 的 9 个数据集上的分类精度对比结果. 由表 2 数据可以看出, 9 个数据集中, 本文方法在其中的 8 个数据集上均取得了最佳结果. 特别是在 Cricket、FingerMovements、HandMovementDirection、Self-RegulationSCP2、StandWalkJump 和 UWaveGestureLibrary 等时间序列较长或维度较多的数据集上优于所有对比方法. 在 Heartbeat 这个时间序列维度较多的数据集上, 本文方法只逊于 WEASEL+MUSE 方法, 优于其余对比算法. 同时, 在平均准确率指标上, 在所有对比方法中最优的为 ShapeNet, 其平均准确率为 66.6%, 而本文方法达到了 73.4%, 相较于对比方法最优结果高了 6.8 个百分点. 进一步验证了本文方法在挖掘维度依赖关系和长期时间依赖关系方面的有效性. 总体而言, 本文方法在挖掘时间序列表征中维度依赖和长程依赖方面有效.

表 2 对比实验结果

数据集	WEASEL+MUSE	ShapeNet	DTWI	DTWD	MLSTM-FCNs	TapNet	MF_NET	Ours
AtrialFibrillation	0.2	0.4	0.133	0.2	0.267	0.2	0.4	<b>0.467</b>
BasicMotions	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	0.9	<b>1.0</b>	<b>1.0</b>	0.825	<b>1.0</b>
Cricket	<b>1.0</b>	0.985	0.979	<b>1.0</b>	0.986	0.917	0.875	<b>1.0</b>
FingerMovements	0.5	0.58	0.570	0.5	0.55	0.44	0.53	<b>0.62</b>
HandMovementDirection	0.270	0.284	0.311	0.203	0.338	0.419	0.568	<b>0.649</b>
Heartbeat	<b>0.776</b>	0.737	0.649	0.678	0.712	0.615	0.692	0.746
SelfRegulationSCP2	0.594	0.578	0.533	0.539	0.528	0.489	0.533	<b>0.616</b>
StandWalkJump	0.4	0.533	0.333	0.2	0.467	0.2	0.4	<b>0.6</b>
UWaveGestureLibrary	<b>0.909</b>	0.899	0.869	0.903	0.872	0.884	0.862	<b>0.909</b>
平均准确率	0.628	0.666	0.597	0.556	0.636	0.574	0.632	<b>0.734</b>

### 3.5 消融实验

#### 3.5.1 模块消融研究

为了研究 3 个模块对提取时间序列表征的影响,设计消融实验来验证. 其中 Ours-nL 表示去除了长期信息表征模块, Ours-nC 表示去除了维度信息表征模块, Ours-I 表示将维度信息表征模块替换为采用式 (1) 维度无关方法获得投影特征. Ours-FC 表示将特征融合层替换为连接操作.

实验结果如表 3 所示, 从表中数据可以看出, 当去除了长期信息表征模块和维度信息表征模块后, 准确率均有所下降. 这表明, 在模型设计中, 既需要考虑维度之间的依赖关系, 也需要考虑长程依赖关系. 进一步地, 将维度信息表征模块替换为按照式 (1) 多维无关形态元变换方法, 实验结果显示, 维度信息表征模块在挖掘维度依赖关系方面具有显著效果. 通过对比 Ours-FC 和 Ours 的实验结果, 发现本文模型的表现优于 Ours-FC, 验证了特征融合层在特征整合中的有效性. 通过消融实验可知, 维度信息表征模块、长期信息表征模块和特征融合层均对于多维时间序列表征起到作用.

表 3 模块消融实验结果

数据集	Ours-nL	Ours-I	Ours-nC	Ours-FC	Ours
AtrialFibrillation	0.4	<b>0.467</b>	0.4	<b>0.467</b>	<b>0.467</b>
BasicMotions	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>
Cricket	0.986	0.958	0.819	0.986	<b>1.0</b>
FingerMovements	0.6	0.59	0.53	0.6	<b>0.62</b>
HandMovementDirection	0.418	0.405	0.378	0.432	<b>0.649</b>
Heartbeat	0.741	0.732	0.689	<b>0.746</b>	<b>0.746</b>
SelfRegulationSCP2	0.572	0.594	0.594	0.594	<b>0.616</b>
StandWalkJump	0.467	0.4	0.533	<b>0.6</b>	<b>0.6</b>
UWaveGestureLibrary	0.806	0.7	0.738	0.875	<b>0.909</b>
平均准确率	0.666	0.650	0.631	0.7	<b>0.734</b>

#### 3.5.2 损失函数消融研究

为了评估损失函数设计的有效性, 本文进行了消融实验, 分别考察基准损失、维度中心损失、长程中心损失和联合损失对模型性能的贡献. 损失函数的 4 种设计如下: 基准损失仅使用负对数似然损失, 为传统分类目标函数, 不包含任何依赖建模; 维度中心损失在基准损失的基础上加入维度依赖特征的中心损失, 用于量化不同维度间的关系; 长程中心损失通过引入时间依赖特征的中心损失, 捕捉时间序列上的长程依赖; 联合损失则同时结合维度中心损失和长程中心损失, 与基准损失一同优化模型特征, 全面考虑维度和时间上的复杂依赖关系. 实验结果见表 4.

表 4 损失函数对比准确率结果

数据集	基准	维度中心损失	长程中心损失	联合损失
AtrialFibrillation	0.333	0.333	0.4	<b>0.467</b>
BasicMotions	0.975	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>
Cricket	0.972	0.986	<b>1.0</b>	<b>1.0</b>
FingerMovements	0.59	<b>0.62</b>	0.60	<b>0.62</b>
HandMovementDirection	0.554	0.635	0.594	<b>0.649</b>
Heartbeat	0.682	0.73	0.72	<b>0.746</b>
SelfRegulationSCP2	0.53	0.54	0.6	<b>0.616</b>
StandWalkJump	0.533	0.566	<b>0.6</b>	<b>0.6</b>
UWaveGestureLibrary	0.863	0.884	0.894	<b>0.909</b>
平均准确率	0.670	0.699	0.712	<b>0.734</b>

实验结果显示, 仅使用基准损失的模型在复杂依赖任务上的性能不足, 尤其是在维度较多或依赖关系强的数据集 (如 HandMovementDirection 和 Heartbeat) 中表现较差, 准确率分别为 55.4% 和 68.2%. 加入维度中心损失后, 模型在大多数数据集上的准确率有显著提升, 例如, 在 HandMovementDirection 上提高到 63.5%, 在 Heartbeat 上从 68.2% 提升到 73%. 这是因为维度中心损失有效捕捉了不同维度间的耦合特性, 增强了模型的特征表达能力. 然而, 在低维数据集 (如 Atrial-Fibrillation) 中, 由于维度间关系简单, 该损失的作用较为有限.

长程中心损失在时间序列较长的数据集上表现尤为显著, 例如, 在 Cricket 上, 准确率从 97.2% 提升到 100%. 在 SelfRegulationSCP2 上, 准确率从 53% 提升到 60%. 这些提升表明, 长程中心损失通过捕捉时间上的全局依赖性, 使模型能够关注关键时间点的信息. 然而, 在序列较短或时间依赖不显著的数据集 (如 Basic-Motions) 中, 其提升相对有限.

联合损失结合了维度和长程中心损失的优势, 显著提升了模型在所有数据集上的性能. 例如, 在 Hand-MovementDirection 上准确率提高至 64.9%; 在 Heartbeat 上准确率达到 74.6%; 同时在 Cricket 和 BasicMotions 中准确率均达到 100% 的最佳表现.

图 2 展示了联合损失的收敛曲线 (我们以 Cricket 数据集为例, 其余数据集也类似, 以训练轮次作为横坐标, 以损失值取对数的结果作为纵坐标, 从而更好展示结果), 收敛曲线先快速下降然后再缓慢下降, 最后趋于平稳, 由此可见联合损失能够达到收敛.

实验结果表明, 联合损失能够有效平衡分类目标和特征学习目标, 充分捕捉维度和时间上的复杂依赖特性, 从而显著提升模型对多维时间序列的表征能力, 且联合损失能够收敛. 这验证了损失函数设计的合理

性以及不同损失模块的协同作用。

### 3.6 注意力研究

#### 3.6.1 维度注意力

为验证注意力机制在维度信息表征中的作用, 本文对维度注意力矩阵  $A_{dim}$  进行了可视化分析, 以揭示其在捕捉多维时间序列中维度间耦合关系方面的效果. 如图 3 所示, 热力图展示了 9 个数据集中不同维度之间的相关性, 颜色越深表示相关性越强. 通过对比不同数据集的注意力分布, 可以直观了解注意力机制如何分配权重并提取关键维度特征.

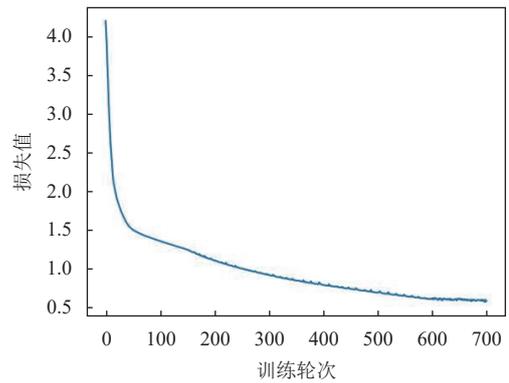


图 2 收敛曲线图

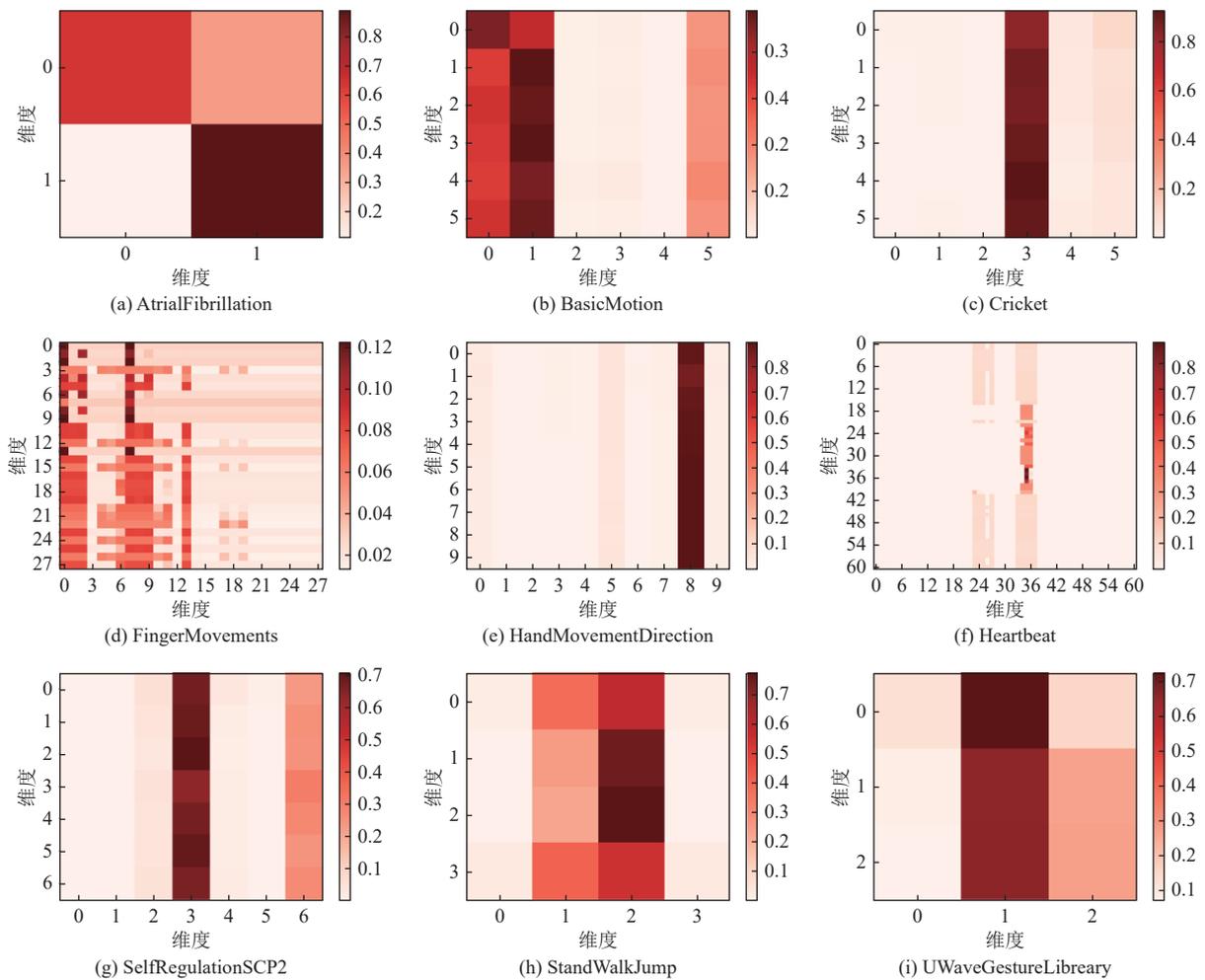


图 3 维度注意力热力图

以 FingerMovements 数据集为例, 该数据集具有 28 个维度, 其中注意力矩阵显示维度 7 的相关性最高, 与其他维度的权重分布差异显著. 这表明维度 7 包含了最重要的信息, 因此被赋予更高的权重. 通过为关键维度分配更多权重, 注意力机制有效忽略了冗余信息,

从而减少了维度之间的干扰. 这种针对性增强特性使得模型能够更准确地表征数据, 并提升分类性能.

此外, 注意力机制的动态加权特性在高维数据集 (如 Heartbeat 和 HandMovementDirection) 中尤为重要. 例如, 在 Heartbeat 中, 维度间的相关性差异较大, 而注

注意力矩阵能够根据每个维度的重要性动态调整权重,使关键维度信息更突出.而在维度较少的数据集(如AtrialFibrillation)中,注意力机制的作用相对较小,因为低维度数据的依赖关系较为简单.

综合来看,注意力机制不仅能够有效捕捉多维时间序列中维度之间的复杂耦合关系,还能够通过动态调整权重分配增强特征表达的精确性.热力图的可视化结果验证了维度注意力机制在忽略冗余维度、突出

关键信息方面的显著作用,从而进一步提升了模型对复杂多维时间序列的表征能力.

### 3.6.2 长程注意力

为验证注意力机制在长程信息表征中的作用,本文对时序注意力得到的相似性权重  $A_{long}$  进行了可视化分析,以揭示其在捕捉时间序列中长程依赖关系方面的效果.如图4所示,热力图展示了9个数据集中不同时间片之间的相关性,颜色越深表示时间片之间相关性越强.

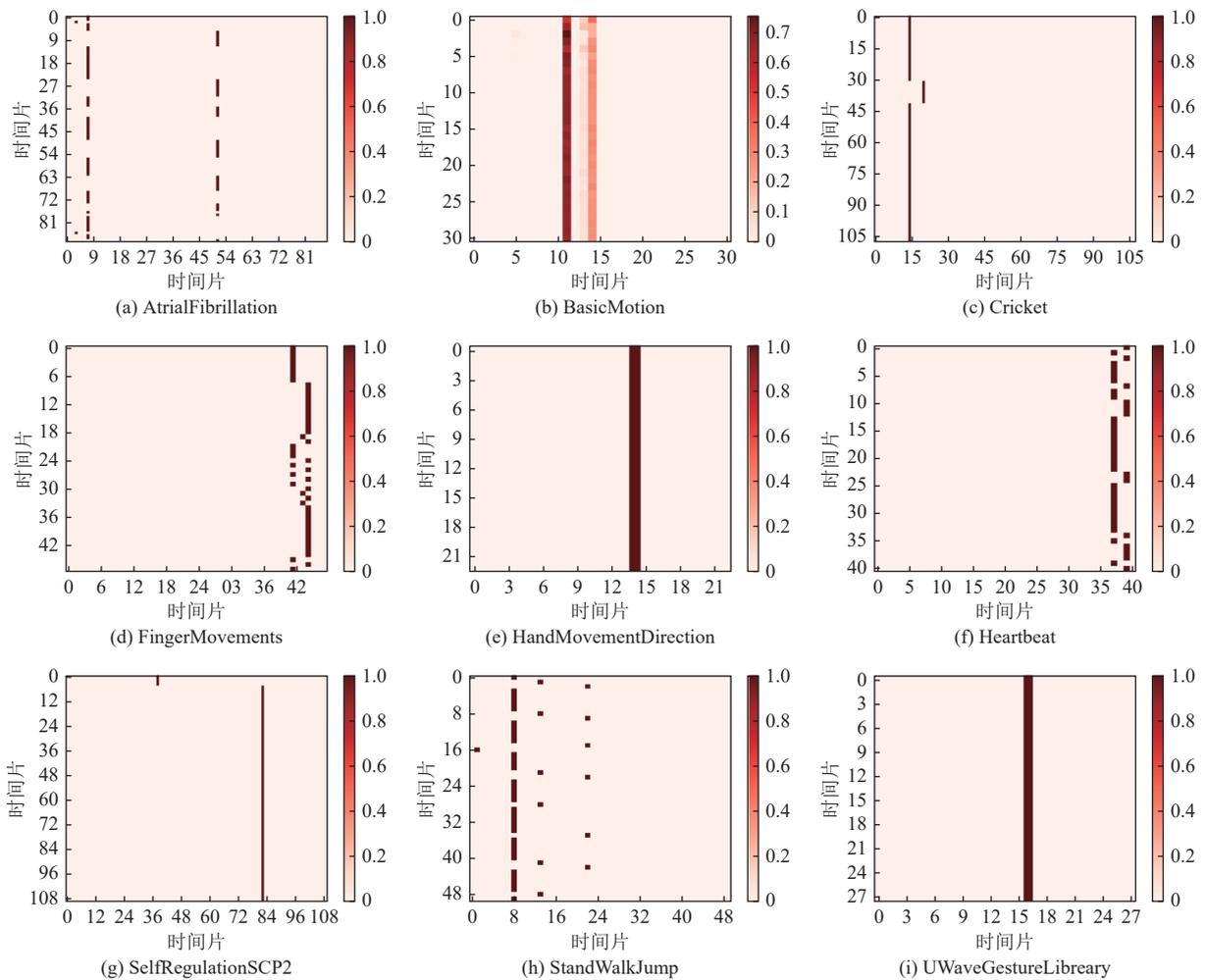


图4 长程注意力热力图

以AtrialFibrillation数据集为例,该数据集有2个维度且长度为640个时间点,形态元长度为32,每次滑动5个时间步,可以得到87个时间片,其中第7个时间片和第51个时间片信息相关性高,与其他时间片的权重分布差异显著.这表明这两个时间片包含了最重要的信息,因此被赋予更高的权重.时序注意力机制捕获了早期(第7个时间片)和中期(第51个时间片)的

依赖关系,对重要时间的信息特征进行增强,这种针对性增强特性使得模型能够更准确地表征数据,并提升分类性能.

时序注意力机制有效捕捉多维时间序列中长程依赖关系,通过对重要时间信息特征增强表达,从而提高精确性.热力图的可视化结果验证了时序注意力机制在捕获长程依赖方面的显著作用,从而进一步提升了

模型对复杂多维时间序列的表征能力

## 4 结论与展望

针对多维时间序列表征中形态元变换未能充分考虑维度间依赖关系和长程依赖关系的问题,提出了一种耦合了维度依赖关系和长程依赖关系的表征方法,进一步挖掘多维时间序列的特征,并通过实验验证了其有效性。首先,本文设计了维度信息表征模块,旨在捕捉多维时间序列数据中各维度之间的复杂交互与依赖关系,生成了富含维度间依赖信息的全新空间投影特征。其次,长期信息表征模块通过融合时序注意力机制与 LSTM,专注于挖掘时间序列数据中长程依赖关系,提升了模型对长期时间依赖模式的识别能力。此外,构建了一个特征融合层,该层将卷积操作与注意力机制相结合,对维度信息表征模块和长期信息表征模块输出的特征进行深度融合与优化。这不仅有效提炼了特征表达,还增强了模型在复杂时间序列模式下的识别能力。在多个数据集上的实验验证表明,本文提出的方法展现出了卓越的性能,充分表明了其在提取多维时间序列表征问题上的有效性与先进性。

### 参考文献

- 1 朱海浩,祝永新,汪辉.基于深度置信网络的多变量时间序列分类方法.计算机仿真,2021,38(12):262-266.[doi:10.3969/j.issn.1006-9348.2021.12.053]
- 2 Ismail Fawaz H, Forestier G, Weber J, *et al.* Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, 2019, 33(4): 917-963. [doi: 10.1007/s10618-019-00619-1]
- 3 Han M, Xu ML. Laplacian echo state network for multivariate time series prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(1): 238-244. [doi: 10.1109/TNNLS.2016.2574963]
- 4 罗伟.基于时间序列表征的检索与分类研究[博士学位论文].济南:山东大学,2022.
- 5 魏池璇,王志海,原继东,等.时间序列可变尺度的时频特征求解及其分类.软件学报,2022,33(12):4411-4428.[doi:10.13328/j.cnki.jos.006346]
- 6 Middlehurst M, Schäfer P, Bagnall A. Bake off redux: A review and experimental evaluation of recent time series classification algorithms. *Data Mining and Knowledge Discovery*, 2024, 38(4): 1958-2031. [doi: 10.1007/s10618-024-01022-1]
- 7 康文轩,陈黎飞,郭躬德.运动序列的时空结构特征表示模型.智能系统学报,2023,18(2):240-250.[doi:10.11992/tis.202203011]
- 8 Kim SW, Park DH, Lee HG. Efficient processing of subsequence matching with the Euclidean metric in time-series databases. *Information Processing Letters*, 2004, 90(5): 253-260. [doi: 10.1016/j.ipl.2004.02.014]
- 9 Shokoohi-Yekta M, Wang J, Keogh E. On the non-trivial generalization of dynamic time warping to the multi-dimensional case. *Proceedings of the 2015 SIAM International Conference on Data Mining*. Vancouver: Society for Industrial and Applied Mathematics, 2015. 289-297.
- 10 Dvornik M, Hadji I, Derpanis KG, *et al.* Drop-DTW: Aligning common signal between sequences while dropping outliers. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. 2021. 13782-13793.
- 11 Shen DS, Chi M. TC-DTW: Accelerating multivariate dynamic time warping through triangle inequality and point clustering. *Information Sciences*, 2023, 621: 611-626. [doi: 10.1016/j.ins.2022.11.082]
- 12 Zheng Y, Liu Q, Chen EH, *et al.* Time series classification using multi-channels deep convolutional neural networks. *Proceedings of the 15th International Conference on Web-age Information Management*. Macao: Springer, 2014. 298-310.
- 13 Karim F, Majumdar S, Darabi H, *et al.* Multivariate LSTM-FCNs for time series classification. *Neural Networks*, 2019, 116: 237-245. [doi: 10.1016/j.neunet.2019.04.014]
- 14 Zhang XC, Gao YF, Lin J, *et al.* TapNet: Multivariate time series classification with attentional prototypical network. *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York: AAAI, 2020. 6845-6852.
- 15 Schäfer P, Leser U. Multivariate time series classification with WEASEL+MUSE. arXiv:1711.11343, 2017.
- 16 Ye LX, Keogh E. Time series shapelets: A new primitive for data mining. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge discovery and Data Mining*. Paris: ACM, 2009. 947-956.
- 17 Lines J, Davis LM, Hills J, *et al.* A shapelet transform for time series classification. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Beijing: ACM, 2012. 289-297.
- 18 Cetin MS, Mueen A, Calhoun VD. Shapelet ensemble for multi-dimensional time series. *Proceedings of the 2015 SIAM International Conference on Data Mining*. Vancouver:

- Society for Industrial and Applied Mathematics, 2015. 307–315.
- 19 Medico R, Ruysinck J, Deschrijver D, *et al.* Learning multivariate shapelets with multi-layer neural networks for interpretable time-series classification. *Advances in Data Analysis and Classification*, 2021, 15(4): 911–936. [doi: [10.1007/s11634-021-00437-8](https://doi.org/10.1007/s11634-021-00437-8)]
- 20 Ghalwash MF, Radosavljevic V, Obradovic Z. Extraction of interpretable multivariate patterns for early diagnostics. *Proceedings of the 13th International Conference on Data Mining*. Dallas: IEEE, 2013. 201–210.
- 21 Patri OP, Sharma AB, Chen HF, *et al.* Extracting discriminative shapelets from heterogeneous sensor data. *Proceedings of the 2014 IEEE International Conference on Big Data*. Washington: IEEE, 2014. 1095–1104.
- 22 Li GZ, Choi B, Xu JL, *et al.* ShapeNet: A shapelet-neural network approach for multivariate time series classification. *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. AAAI, 2021. 8375–8383.
- 23 Bostrom A, Bagnall A. A shapelet transform for multivariate time series classification. arXiv:1712.06428, 2017.
- 24 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)]
- 25 许珠路, 王兴芬, 刘亚辉. 融合 CNN-BiLSTM-Attention 的集成学习价格预测. *计算机系统应用*, 2023, 32(6): 32–41. [doi: [10.15888/j.cnki.csa.009137](https://doi.org/10.15888/j.cnki.csa.009137)]
- 26 Bagnall A, Dau HA, Lines J, *et al.* The UEA multivariate time series classification archive, 2018. arXiv:1811.00075, 2018.
- 27 Du MS, Wei YX, Zheng XW, *et al.* Multi-feature based network for multivariate time series classification. *Information Sciences*, 2023, 639: 119009. [doi: [10.1016/j.ins.2023.119009](https://doi.org/10.1016/j.ins.2023.119009)]

(校对责编: 王欣欣)