Statistics and Application 统计学与应用, 2026, 15(4), 112-126
Published Online April 2026 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2026.154076

Hans汉斯

面向长序列数据分类的变长子序列最优传输模型

胡佳美，李文静，陈继强*

河北工程大学数理科学与工程学院，河北 邯郸

收稿日期：2026年3月10日；录用日期：2026年3月30日；发布日期：2026年4月14日

摘　要

序列数据分类是数据挖掘领域的重要任务，其性能高度依赖于序列间相似性度量的有效性。动态时间规整及其变体作为当前广泛采用的相似性度量方法，在长序列数据分类中存在计算复杂度高、病态对齐、忽略局部结构特征以及过度最大化相似性等问题。为此，本文提出一种面向长序列数据分类的变长子序列最优传输模型。首先，首次将感知重要点与最优传输理论结合，构造变长子序列，以子序列匹配替代传统逐点匹配，显著降低计算复杂度并缓解病态对齐。其次，为加强序列局部结构特征，引入复杂度修正因子构建代价矩阵。再次，设计位置惩罚正则项，以抑制位置过远子序列的不合理匹配，并缓解了过度最大化相似性问题。最后，采用熵正则化最优传输框架，通过Sinkhorn-Knopp快速迭代算法实现高效求解。在20个UCR数据集上的实验结果表明，所提模型在分类准确率和计算效率方面均显著优于主流对比方法，验证了其在长序列数据分类任务中的有效性与可行性。

关键词

序列数据分类，最优传输，子序列匹配，相似性度量

Variable-Length Subsequence Optimal Transport Model for Long Sequence Data Classification

Jiamei Hu, Wenjing Li, Jiqiang Chen*

School of Mathematics and Physics, Hebei University of Engineering, Handan Hebei

Received: March 10, 2026; accepted: March 30, 2026; published: April 14, 2026

*通讯作者。

文章引用: 胡佳美, 李文静, 陈继强. 面向长序列数据分类的变长子序列最优传输模型[J]. 统计学与应用, 2026, 15(4): 112-126. DOI: 10.12677/sa.2026.154076

Abstract

Sequence data classification is a fundamental task in data mining, and its performance heavily depends on the effectiveness of similarity measures between sequences. Dynamic Time Warping (DTW) and its variants, as widely adopted similarity measures, suffer from high computational complexity, pathological alignments, neglect of local structural characteristics, and excessive similarity amplification when applied to long sequence classification. To address these issues, this paper proposes a variable-length subsequence optimal transport model for long sequence data classification. First, by integrating Perceptually Important Points (PIPs) with optimal transport theory, we construct variable-length subsequences to replace traditional point-wise matching with subsequence-level matching, significantly reducing computational complexity while mitigating pathological alignments. Second, a complexity correction factor is introduced to construct a cost matrix that captures both numerical differences and local structural characteristics between subsequences. Third, a position penalty regularization term is designed to suppress unreasonable matches between subsequences that are far apart in temporal position, effectively alleviating excessive similarity amplification. Finally, the model is formulated within an entropy-regularized optimal transport framework and efficiently solved using the Sinkhorn-Knopp iterative algorithm. Experimental results on 20 UCR datasets demonstrate that the proposed model significantly outperforms mainstream methods in both classification accuracy and computational efficiency, validating its effectiveness and feasibility for long sequence data classification tasks.

Keywords

Sequence Data Classification, Optimal Transport, Subsequence Matching, Similarity Measure

1. 引言

序列数据分类是时间序列数据挖掘中的核心任务，广泛应用于医疗监测[1]、金融分析[2]和工业传感[3]等领域。该任务的关键在于设计有效的相似性度量方法以实现可靠的分类。动态时间规整(Dynamic Time Warping, DTW) [4]及其变体[5]-[13]作为序列数据中最精确的相似性度量方法之一，通过非线性映射来对齐序列，但其“一对多”匹配易产生病态对齐，并且计算复杂度与序列长度成正比，导致长序列数据分类准确性和计算效率大幅下降[7]。

近年来，为克服 DTW 的局限性，研究者们在以下三个方面展开了探索。首先，在序列对齐方面[9]-[12]。Hao 等[9]和 Hong 等[10]利用离散小波变换提取局部结构信息，通过加权机制融合序列数值差异和局部形态结构，提升了对齐路径的合理性。Li 等[11]通过插值将多维信号映射到一维时间特征，并结合受限路径搜索，实现了多传感器数据的时空对齐。而 Qiu 等[12]通过考虑局部极值的位置信息，避免了因极值引发的序列对齐问题。其次，计算效率方面[13]-[18]。金俊超[13]等通过重排序、级联下界约束、全局约束和提前抛弃等机制，有效减少了不必要的 DTW 计算。Haviana 等[16]采用多级粗化、投影和细化策略，先递归压缩序列并在低分辨率计算规整路径，再投影至高分辨率局部优化，显著降低时间复杂度。最后，基于子序列的方法[19]-[27]。Ye 等[19]开创性地提出了 Shapelet 概念，通过有监督地搜索最具判别力的子序列进行分类。Mishra 等[24]将序列分割为子序列，以实现对具有同步形态特征序列的相似性度

量。Zhang 等[27]提出了基于感知重要点(Perceptually Important Points, PIPs)的 PIPDTW 算法，用于高效进行传感器数据的 Top-k DTW 相似性搜索。PIPs 的核心优势在于，将冗长的原始数据压缩为保留核心形态特征的紧凑表示，这不仅极大地减少了数据处理的规模，还显著提升了相似性搜索的效率。

尽管上述方法取得了一定进展，但在处理复杂长序列时仍面临关键挑战。一方面，DTW 及其改进的方法虽能提升匹配的合理性，但其严格单调对齐机制难以避免病态对齐，且计算复杂度随序列长度急剧增加。另一方面，基于子序列的方法多依赖固定窗口或启发式分割，在自适应捕捉序列的多尺度局部特征方面存在局限。针对上述问题，最优传输(Optimal Transport, OT) [28]框架为序列相似性度量提供了新的思路。与传统对齐方法不同，OT 将序列视为概率分布，通过灵活的“多对多”匹配机制，缓解了 DTW 病态对齐的问题。Zheng 等[29]通过将序列表示为包含观测值与时间坐标的二维时空点，实现数值与时间的联合对齐。Su 等[30]通过引入逆差分矩作为顺序对齐正则项以保持时序一致性，同时提高对齐合理性。Latorre 等[31]将 Soft-DTW 与正则化 OT 相结合，在可微框架下同时实现序列数据的对齐和空间差异的度量，并导出闭式解以提高计算效率。尽管这些方法推动了 OT 在序列分析中的初步应用，但现有 OT 方法大多将序列建模为离散观测点的分布，仍缺乏对序列内部局部结构信息的有效利用。

PIPs 方法[32]通过递归识别序列数据中的关键转折点与极值点，能够自适应构建具有明确语义的局部子序列，为弥补 OT 方法的不足提供了可能。与传统需要预设窗口的固定子序列长度方法不同，PIPs 方法可自然形成具有明确语义的变长子序列，为 OT 匹配提供更具结构性的单元。然而，现有方法在计算子序列之间的距离时往往忽略了子序列在整条序列中的位置，仅关注形状是否相似，而不考虑是否出现在相近的位置。这会导致两条本质上不同的序列数据，如果刚好在不同位置包含相似形状的子序列，则会被错误地判定为高度相似，从而过度放大了它们的整体相似度。

为此，本文提出一种面向长序列数据分类的变长子序列最优传输模型(PISOT)。主要创新点如下：

(1) 首次将 PIPs 与 OT 理论结合。实现从“逐点匹配”到“子序列匹配”的转变，在保留序列数据关键形态特征的同时大幅提升计算效率。

(2) 设计代价矩阵。采用“短片段在长片段上滑动”的策略寻找最优局部对齐窗口，同时引入复杂度不变性校正因子，使传输代价同时反映子序列间的数值差异与形态波动特征。

(3) 设计位置惩罚正则项。抑制位置相距过远的子序列误匹配，并有效缓解过度最大化相似性问题。

(4) 在 UCR 序列数据集上进行实验。验证了 PISOT 模型分类准确性与计算效率方面的优越性。

本文结构如下。第 2 节介绍预备知识。第 3 节构建变长子序列最优传输模型(PISOT)。第 4 节通过实验验证模型的有效性。第 5 节为结论。

2. 预备知识

2.1. Wasserstein 距离

Wasserstein 距离是一种衡量两个概率分布之间差异的方法，广泛应用于数据挖掘和机器学习任务中。设定两个离散点集 $X=\{x_i\}_{i=1}^{N}$ 和 $Y=\{y_j\}_{j=1}^{M}$，其对应的概率分别表示为 $\mu_X=\sum_{i=1}^{N}a_i\delta_{x_i}$ 和 $\nu_Y=\sum_{j=1}^{M}b_j\delta_{y_j}$，其中 a_i 和 b_j 分别表示序列 X 和 Y 中离散点的权重。当缺乏先验信息时，通常将权重设置为均匀分布，即 $a_i=\frac{1}{N}$ 和 $b_j=\frac{1}{M}$。Wasserstein 距离的目标是通过寻找一个最优的传输矩阵 T，来最小化两个分布之间的距离。其数学表达式为：

$$\min_{T\in\pi(a,b)}\langle D,T\rangle_F=\min_{T\in\pi(a,b)}\sum_{i=1}^{N}\sum_{j=1}^{M}d_{i,j}t_{i,j} \tag{1}$$

$$\pi(a,b)=\left\{T\in\mathbb{R}_{+}^{N\times M}\,\middle|\,T1_M=a,T1_N=b\right\} \tag{2}$$

其中 $d_{i,j}$ 表示从数据点 x_i 到数据点 y_j 的传输成本， $t_{i,j}$ 表示从 x_i 传输到 y_j 的质量。

2.2. 感知重要点

感知重要点(PIPs)最初由 Chung 等[32]提出，是一种经典的序列数据稀疏表示方法。该方法通过选取序列中具有结构意义的极值点或转折点，以少量代表性重要点近似原始序列的整体形态，实现数据压缩与特征保留的平衡。

对于长度为 N 的序列数据 $X=\{x_1,x_2,\cdots,x_N\}$，PIPs 的提取过程如图 1 所示：首先将序列的首尾点加入到 PIPs 列表中，记为 $\mathcal{P}=[1,N]$。然后采用贪心策略，从剩余点中依次选取对当前折线近似误差贡献最大的点加入集合，直至 $|\mathcal{P}|=k\quad(k<<N)$为止。

设当前已选的 PIPs 索引集合按升序排列为：

$$\mathcal{P}=\{p_1,p_2,\cdots,p_k\},\quad 1=p_1<p_2<\cdots<p_k=N \tag{3}$$

计算每个候选点 $z(z\notin\mathcal{P})$ 到相邻 PIPs 所构成线段的垂直距离：

$$PD(z;\mathcal{P})=\frac{\left|e\cdot(z-p_g)-x_z+x_{p_g}\right|}{\sqrt{e^2+1}} \tag{4}$$

其中 $p_g<z<p_{g+1}$， $e=\dfrac{x_{p_{g+1}}-x_{p_g}}{p_{g+1}-p_g}$。

选取使垂直距离最大的点加入 $\mathcal{P}$ 集合：

$$\mathbf{z}^{*}=\underset{z\in\{2,\cdots,N-1\}}{\arg\max}\,PD(z;\mathcal{P}) \tag{5}$$

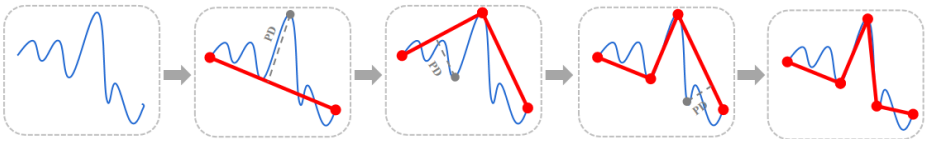

Figure 1. The procedure for extracting the first 5 PIPs
图 1. 提取前 5 个 PIPs 的过程

3. 变长子序列最优传输模型

3.1. 变长子序列

传统的子序列提取方法多采用固定长度滑动窗口，难以适应序列在时间尺度上的动态变化。窗口过小会丢失全局结构，窗口过大则易导致局部特征失真。为此，本文引入 PIPs 构建变长子序列，使子序列长度能够依据数据自身的关键转折点自适应调整。

为消除幅值偏移与缩放对相似性度量的影响，首先对原始序列进行 z-score 标准化。给定两条序列数据 $X=(x_1,\cdots,x_N)\in\mathbb{R}^N$ 与 $Y=(y_1,\cdots,y_M)\in\mathbb{R}^M$，标准化过程定义为：

$$\begin{aligned}\tilde{X}_N&=\frac{x_n-\mu_X}{\sigma_X+\omega},\quad n=1,\cdots,N\\ \tilde{Y}_N&=\frac{y_m-\mu_Y}{\sigma_Y+\omega},\quad m=1,\cdots,M\end{aligned} \tag{6}$$

其中 μ_X,σ_X 和 μ_Y,σ_Y 分别为序列 X 与 Y 的均值与标准差，$\omega>0$。为简化记号，下文仍以 X,Y 表示标准化后的序列。

利用(5)式从序列 X 和序列 Y 中分别提取 k 个 PIPs，记为：

$$\begin{aligned}\mathcal{P}^x&=\left\{p_1^x,\cdots,p_k^x\right\},\quad 1=p_1^x<\cdots<p_k^x=N\\ \mathcal{P}^y&=\left\{p_1^y,\cdots,p_k^y\right\},\quad 1=p_1^y<\cdots<p_k^y=M\end{aligned}\tag{7}$$

为保留序列数据局部结构并维持上下文信息，采用连续 3 个 PIPs 构建变长子序列。子序列的数量为 $k-2$，定义如下：

$$\begin{aligned}S_i^x&=X\left[p_i^x;p_{i+2}^x\right],\quad i=1,\cdots,k-2\\ S_j^y&=Y\left[p_j^y;p_{j+2}^y\right],\quad j=1,\cdots,k-2\end{aligned}\tag{8}$$

由于 PIPs 间隔不均匀，子序列的长度可以变化，其长度分别为：

$$\begin{aligned}L_i^x&=p_{i+2}^x-p_i^x+1,\quad i=1,\cdots,k-2\\ L_j^y&=p_{j+2}^y-p_j^y+1,\quad j=1,\cdots,k-2\end{aligned}\tag{9}$$

3.2. 代价矩阵

由于 PIPs 分布的非均匀性，3 个连续 PIPs 构建的子序列长度不同。若直接计算两子序列间的逐点距离，会产生长度错位与边界失配问题。为解决此问题，提出基于滑动匹配的子序列距离度量策略，将较短的子序列在较长的子序列上进行滑动对齐，通过寻找最佳匹配位置来缓解边界差异带来的影响。

给定任意子序列对 $\left(S_i^x,S_j^y\right)$，设其长度满足 $L_i^x\le L_j^y$。在较长的子序列 S_j^y 上滑动长度为 L_i^x 的窗口，生成候选对匹配片段：

$$S_j^y(r)=S_j^y\left[r:r+L_i^x-1\right],\quad r=1,\cdots,L_j^y-L_i^x+1\tag{10}$$

其中 $S_j^y(r)$ 表示从 $S_j^y(r)$ 的第 r 个元素开始、长度为 L_i^x 的子片段，计算其与 S_i^x 的等长距离。

对任意等长序列 $U=(u_1,\cdots,u_L)$ 和 $V=(v_1,\cdots,v_L)$，定义均方根距离为：

$$d_{rms}(U,V)=\sqrt{\frac{1}{L}\sum_{l=1}^{L}(U_l-V_l)^2}\tag{11}$$

仅用 d_{rms} 仍可能无法有效区分数值相近但波动形态差异的片段。例如：一个高复杂度片段与一个更平滑的片段可能得到近似的 d_{rms}，但二者的“形态复杂度”差异显著。为此，引入复杂度不变性[33]来量化序列的形态复杂度：

$$CI(U)=\sqrt{\sum_{l=1}^{L-1}(u_{l+1}-u_l)^2+\omega}\tag{12}$$

进而定义复杂度修正因子[33]为：

$$\rho(U,V)=\frac{\max\left\{CI(U),CI(V)\right\}}{\min\left\{CI(U),CI(V)\right\}}\tag{13}$$

当两个子序列复杂度差异显著时，$\rho(U,V)>1$ 将放大(11)式距离，从而抑制数值相近但序列波动形态差异大的错误匹配。

综合滑动对齐与复杂度修正因子，定义子序列对 $\left(S_i^x,S_j^y\right)$ 距离为：

$$d_{CI\_seg}\left(S_i^x,S_j^y\right)=\begin{cases}\min\limits_r\left(d_{rms}\left(S_i^x,S_j^y(r)\right)*\rho\left(S_i^x,S_j^y(r)\right)\right), & if \quad L_i^x\le L_j^y\\ \min\limits_r\left(d_{rms}\left(S_j^y,S_i^x(r)\right)*\rho\left(S_j^y,S_i^x(r)\right)\right), & if \quad L_i^x>L_j^y\end{cases} \tag{14}$$

该定义通过遍历所有可能的滑动位置 r 寻找最佳局部对齐，并引入复杂度修正因子对距离进行调整。基于此，构建代价矩阵：

$$D_{CI\_seg}(i,j)=d_{CI\_seg}\left(S_i^x,S_j^y\right),\quad i=1,\cdots,k-2\quad j=1,\cdots,k-2 \tag{15}$$

3.3. 位置惩罚正则项

为抑制最优传输计划中位置相距过远的子序列进行匹配，构建了位置惩罚正则项。计算每对变长子序列的归一化中心位置：

$$\begin{aligned}\tau_i^x&=\frac{p_i^x+p_{i+2}^x}{2(N-1)},\quad i=1,\cdots,k-2\\ \tau_j^y&=\frac{p_j^y+p_{j+2}^y}{2(M-1)},\quad j=1,\cdots,k-2\end{aligned} \tag{16}$$

然后，定义位置惩罚项为：

$$\lambda_{pos}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\tau_i^x-\tau_j^y\right)^2 t_{i,j} \tag{17}$$

其中 $\lambda_{pos}>0$ 为位置惩罚系数，用于控制子序列位置对齐的程度。当 λ_{pos} 取值较大时，模型将倾向于匹配中心位置相近的子序列。

3.4. 模型构建

将序列 X 和序列 Y 对应的变长子序列集合视为两个离散分布的支撑点。为反映不同子序列的信息含量差异，本文根据子序列长度定义最优传输的边际概率分布：

$$\begin{aligned}a_i&=\frac{L_i^x}{\sum_{r=1}^{k-2}L_r^x},\quad i=1,\cdots,k-2\\ b_j&=\frac{L_j^y}{\sum_{s=1}^{k-2}L_s^y},\quad j=1,\cdots,k-2\end{aligned} \tag{18}$$

基于代价矩阵 $D_{CI\_seg}(i,j)$ 与位置惩罚正则项，构建变长重要子序列最优传输模型：

$$T^*=\arg\min_{T\in\pi(a,b)}\left\langle T,D_{CI\_seg}\right\rangle_F+\lambda_{pos}\sum_{i=1}^{k-2}\sum_{j=1}^{k-2}\left(\tau_i^x-\tau_j^y\right)^2 t_{i,j}+\varepsilon H(T) \tag{19}$$

$$\pi(a,b)=\left\{T\in\mathbb{R}_+^{(k-2)\times(k-2)}\,\middle|\,T1_{k-2}=a,T^T1_{k-2}=b\right\} \tag{20}$$

其中 $H(T)=\sum_{i=1}^{k-2}\sum_{j=1}^{k-2}t_{i,j}\left(\log t_{i,j}-1\right)$，基于上面的模型，定义两条序列之间的相似性度量为：

$$PISOT(X,Y)=\left\langle T^*,D_{CI\_seg}\right\rangle_F \tag{21}$$

3.5. 算法求解

模型(19)式的求解可通过 Sinkhorn-Knopp 算法高效实现。该算法基于矩阵缩放思想，通过迭代更新缩放因子使传输矩阵逐渐满足边际约束，求出最优传输矩阵 T^*，代入(21)式计算序列间的相似性度量，

然后对序列数据进行分类。具体实现步骤如算法 1 所示。

算法 1：基于变长子序列最优传输模型的序列数据分类算法

输入：标准化后的训练集 $\mathcal{X}_{train}=\{X_1,\cdots,X_n\}$ 和测试集 $\mathcal{Y}_{test}=\{Y_1,\cdots,Y_m\}$，及训练集标签 $[c_1,\cdots,c_n]$，PIPs 数量 k，常数 $\omega>0$，正则化系数 λ_{pos} 和 ε，最大迭代次数 $I_{\max}$，收敛阈值 θ。

输出：最优传输矩阵 T^*，相似性度量 $PISOT\left(X_q,Y_w\right)$，测试集数据的预测标签 $[\hat{c}_1,\cdots,\hat{c}_m]$。

1) for 第 w 个测试序列 Y_w do

通过(5)式提取测试集 PIPs $\mathcal{P}^Y\leftarrow \text{PIPSExteactor}\left(Y_w,k\right)$

通过(8)和(9)式构造子序列并计算子序列的长度

通过(16)式计算子序归一化中心位置

通过(18)式计算边概率分布

2) for 第 q 个训练序列 X_q do

通过(5)式提取测试集 PIPs $\mathcal{P}^X\leftarrow \text{PIPSExteactor}\left(X_q,k\right)$

通过(8)和(9)式构造子序列并计算子序列的长度

通过(16)式计算子序归一化中心位置

通过(18)式计算边概率分布

通过(15)式构建代价矩阵，并添加(17)式位置惩罚项构建完整代价矩阵

3) 利用 Sinkhorn-Knopp 迭代求解(19)式的最优传输问题

$K\leftarrow \exp\left(-D/\varepsilon\right)$

$u\leftarrow 1_{k-2},v\leftarrow 1_{k-2}$

计算最优传输矩阵 $T^*\leftarrow diag\left(u\right)Kdiag\left(v\right)$

4) 计算相似性度量 $PISOT(X_q,Y_w)\leftarrow\left\langle T^*,D_{CI\_seg}\right\rangle_F$

end for

5) 计算相似性度量的最小值 $\hat{q}=\arg\min_q PISOT\left(X_q,Y_w\right)$

赋予最近邻样本标签 $\hat{c}_w\leftarrow c_q$

end for

返回预测标签 $[\hat{c}_1,\cdots,\hat{c}_m]$

3.6. 算法复杂度分析

本文提出的 PISOT 模型的算法复杂度主要来源于变长子序列构造、代价矩阵构建以及 Sinkhorn 迭代求解三个关键步骤。假设训练集序列长度为 N，测试集序列长度为 M，$N\geq M$。提取 k 个 PIPs，构造 $m=k-2$ 个变长子序列，其中最大子序列长度为 $L_{\max}$。PIPs 提取采用贪心算法，复杂度为 $O(kN)$。对于每一对训练与测试序列，需要构建 $m\times m$ 的代价矩阵。计算任意子序列对 $\left(S_i^x,S_j^y\right)$ 的滑动复杂度修正距离，需要在其较长子序列上进行滑动窗口匹配，复杂度为 $O\left(L_{\max}^2\right)$，因此构建完整代价矩阵的复杂度为 $O\left(m^2L_{\max}^2\right)$。随后，通过 Sinkhorn 算法求解正则化最优传输问题，每次迭代涉及矩阵与向量乘法，复杂度为 $O\left(m^2\right)$，设收敛所需迭代次数为 I，则迭代阶段总复杂度为 $O\left(Im^2\right)$。获得最优传输计划 T^* 后，计算相似性度量的复杂度为 $O\left(m^2\right)$。因此，对于一个测试样本与一个训练样本的匹配，PISOT 模型的总算法复杂度为 $O\left(m^2\left(L_{\max}^2+I\right)\right)$。

4. 实验

为验证本文提出的 PISOT 模型的有效性，本文在 20 个公开的 UCR 数据集上开展了实验。分别采用

质心分类和 1NN 分类两种策略，从分类准确率和计算效率两方面对模型性能进行综合评估。所有实验均在配备 i7-4800U 和 16GB RAM 的标准 Windows 11 操作系统上完成。

4.1. 数据集描述

本实验从 UCR 数据库[34]中选取了 20 个具有代表性的序列数据集。这些数据集在序列长度(96~4250)、类别数量(2~52)及数据维度(1~10)上分布广泛，能够有效评估模型在不同复杂度场景下的鲁棒性与泛化能力。特别地，为验证 PISOT 模型在处理长序列数据方面的优势，本文重点关注序列长度超过 1000 的数据集[35]，以探究变长子序列构建机制与 OT 方法在提升分类准确率的同时，是否能够显著提高计算效率。所选数据集的详细描述信息如表 1 所示。

Table 1. Dataset description
表 1. 数据集描述

数据编号	数据集	训练集	测试集	长度	维度	类别
Computers	Computers	250	250	760	1	2
EGG2	ECG200	100	100	96	1	2
EOGVS	EOG Vertical Signal	362	362	1250	1	12
FU	Faces UCR	200	2050	131	1	14
HA	Haptics	155	308	1092	1	5
HE	Herring	64	64	512	1	2
HO	Hand Outlines	1000	370	2709	1	2
HT	House Twenty	40	119	2000	1	2
IS	Inline Skate	100	550	1882	1	7
Mallat	Mallat	55	2345	1024	1	8
PAP	Pig Art Pressure	104	208	2000	1	52
Rock	Rock	20	50	2844	1	4
SL	Swedish Leaf	500	625	128	1	15
WS	Word Synonyms	267	638	270	1	25
WO	Worms	181	77	900	1	5
AWR	Articulary Word Recognition	275	300	144	9	25
CMJ	Counter Movement Jump	419	179	4250	3	3
HMD	Hand Movement Direction	160	74	400	10	4
SWJ	Stand Walk Jump	12	15	2500	4	3
SRS2	SelfRegulationSCP2	200	180	1152	7	2

4.2. 实验结果

4.2.1. 对比实验

1. 基于质心分类器的对比实验

为验证本文提出的 PISOT 模型在序列数据分类任务中的有效性，选取 LCSS [21]、MSM [22]、DTW [4]、ShapeDTW [23]、SegDTW [24]、PIPDTW [27]及 MPDist [20]等 7 种方法作为对比，在 20 个 UCR 数

据集上进行分类准确率评估。实验首先在计算效率较高的质心分类器框架下开展。该分类器通过提取训练集中各类别的序列质心，将分类任务简化为测试样本与各类质心的相似度比较，从而显著减少匹配次数，降低整体计算开销，适用于对实时性要求较高的实际场景。

Table 2. Accuracy comparison of eight methods based on centroid classifier (%)
表 2. 质心分类器下的 8 种方法准确率对比(%)

数据编号	LCSS	MSM	DTW	ShapeDTW	SegDTW	PIPDTW	MPDist	PISOT
Computers	22.80	58.40	54.00	54.00	52.80	60.00	35.60	**64.40**
ECG2	61.00	72.00	68.00	68.00	69.00	73.00	60.00	**80.00**
EOGVS	23.48	29.01	22.65	22.10	35.08	29.28	20.99	**35.36**
FU	43.51	78.05	80.73	81.51	**82.59**	72.10	33.51	76.88
HE	53.12	45.31	54.69	57.81	40.62	51.56	48.44	**65.62**
HA	38.31	36.36	27.92	31.82	38.31	35.71	22.08	**38.96**
HO	71.62	72.16	70.27	72.43	81.62	**86.22**	68.64	83.51
HT	84.03	76.47	83.19	**85.71**	78.15	56.30	62.18	**85.71**
IS	22.90	24.36	23.27	23.27	16.00	18.91	16.73	**29.82**
Mallat	54.80	85.63	92.96	92.88	**94.75**	91.98	55.99	90.70
WS	36.99	36.83	37.30	38.71	32.76	33.23	21.63	**40.60**
PAP	42.79	37.46	47.12	41.83	41.34	43.27	67.31	**86.54**
Rock	46.00	34.00	**48.00**	36.00	**48.00**	50.00	48.00	46.00
SL	56.32	66.24	59.36	66.72	60.48	66.08	40.80	**72.64**
WO	31.17	37.66	33.77	38.96	28.57	37.66	32.47	**51.95**
AWR	86.00	72.00	83.00	91.00	87.33	71.33	79.33	**93.00**
HMD	20.27	25.68	33.78	27.03	33.78	**37.84**	25.68	**37.84**
SWJ	46.67	40.00	20.00	33.33	13.33	46.67	26.67	**60.00**
SRS2	48.89	46.11	48.89	43.89	48.89	47.78	53.89	**55.56**
CMJ	30.73	35.75	33.52	35.75	31.28	33.52	35.75	**36.31**

从表 2 可见，在质心分类器下，PISOT 模型在全部 20 个数据集中有 16 个数据集取得了最优分类准确率。特别是在长序列数据集(如 PAP、Rock 和 CMJ)上优势显著，其中在 PAP 数据集上准确率达 86.54%，较第二名 MPDist 的准确率 67.31%提升了 19.23%。在 SWJ 数据集上 PISOT 准确率达到 60.00%，较传统 DTW 准确率的 46.67%提升 13.33%。

在计算效率方面，由图 2 可知，PISOT 在质心分类器框架下的平均运行时间显著低于所有对比方法。这一优势源于两方面：一是基于 PIPs 的变长子序列表示大幅降低了匹配复杂度；二是质心分类器较少的匹配次数进一步减少了计算开销。该结果验证了 PISOT 在保持高分类精度的同时，具备优异的计算效率，适用于对实时性要求较高的实际应用场景。

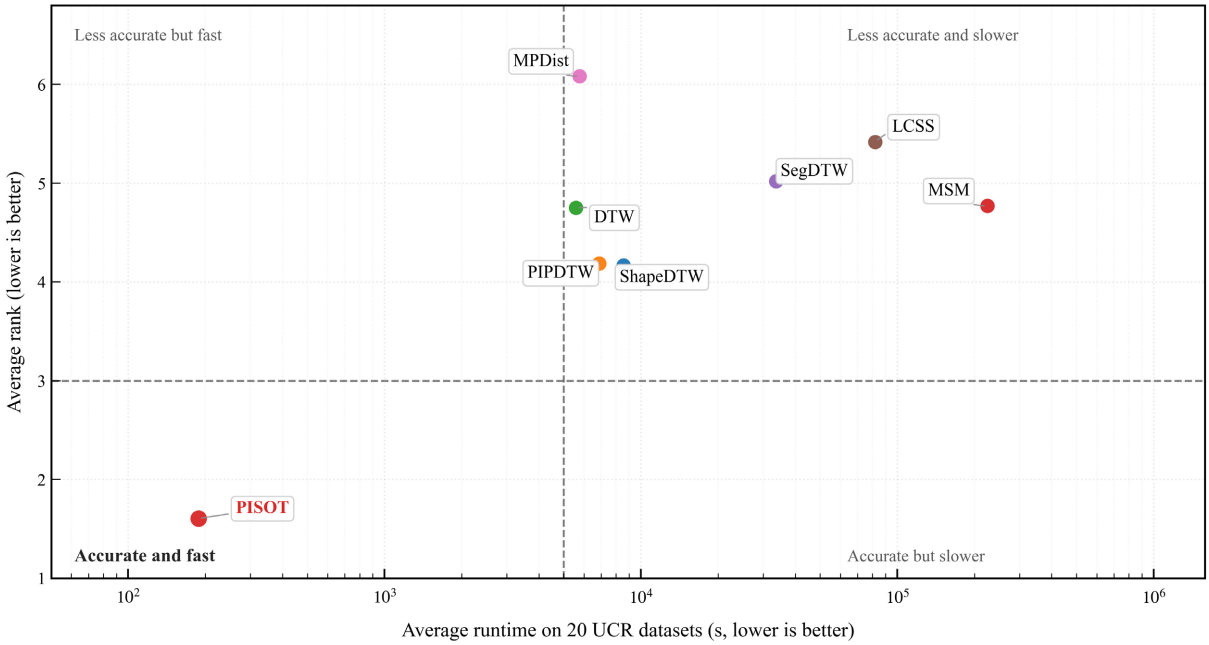

Figure 2. Comparison of average running time and accuracy ranking of 8 methods based on the centroid classifier
图 2. 质心分类器下的 8 种方法平均运行时间和准确率排名对比

2. 基于 1NN 分类器的对比实验

选择 1NN 分类器主要基于两方面考虑：首先，作为一种非参数方法，1NN 直接依赖于距离度量本身，能够更好地反映不同相似性度量的判别能力，避免了复杂分类器参数调整可能引入的偏差。其次，1NN 在序列数据分类领域被广泛视为一种强基准，其结果具有高度的可解释性和可比较性

Table 3. Accuracy comparison of eight methods based on the 1NN classifier (%)
表 3. 1NN 分类器下的 8 种方法准确率对比(%)

数据编号	LCSS	MSM	DTW	ShapeDTW	SegDTW	PIPDTW	MPDist	PISOT
Computers	63.60	58.20	66.80	66.80	55.60	54.80	63.60	**68.00**
ECG2	67.00	84.00	80.00	83.00	80.00	87.00	76.00	**91.00**
EOGVS	43.67	43.10	44.75	44.20	45.58	46.13	43.10	**48.62**
FU	71.61	94.15	93.41	**95.61**	93.56	90.49	50.54	93.51
HA	39.00	37.01	36.36	38.96	36.04	36.69	23.38	**44.16**
HE	46.88	54.69	54.69	51.56	56.25	**60.94**	59.38	56.25
HO	82.43	83.51	80.27	82.43	81.62	**85.14**	79.19	84.86
HT	84.00	91.60	87.39	91.60	85.71	74.79	73.95	**95.80**
IS	36.91	37.27	37.45	36.91	36.00	35.27	18.00	**49.64**
Mallat	87.21	86.82	91.43	92.75	91.94	92.88	60.68	**94.12**
PAP	33.65	31.73	31.25	62.50	36.06	36.06	76.44	**93.75**
Rock	60.00	78.00	64.00	62.00	64.00	80.00	70.00	**82.00**
SL	84.00	87.04	79.04	86.88	79.20	80.48	36.80	**91.52**

续表

WS	71.32	68.18	67.55	69.75	70.69	68.50	21.63	**75.08**
WO	59.74	55.84	51.95	53.25	58.44	54.55	45.45	**68.83**
AWR	19.33	81.67	91.67	96.00	92.67	77.67	93.00	**97.00**
HMD	18.92	29.73	21.62	21.62	25.68	25.68	29.73	**32.43**
SWJ	46.67	46.67	26.67	46.67	40.00	46.67	26.67	**60.00**
SRS2	46.67	55.00	51.67	47.22	48.33	52.22	51.11	**58.33**
CMJ	32.40	34.08	35.75	36.31	32.40	36.49	32.40	**36.87**

从表 3 可以看出，在 1NN 分类器下，PISOT 模型在长序列数据集上优势尤为突出，在 PAP 数据集上，PISOT 准确率达 93.75%，较第二名的 76.44%提升 17.3。此外，在 HT、IS、WS 等数据集上，PISOT 准确率分别达 95.80%、49.64%和 75.08%，均明显优于对比方法。上述结果表明，PISOT 在 1NN 分类器框架下同样具备强大的分类性能，能够有效应对不同长度、不同类别数、不同维度的序列数据分类任务。

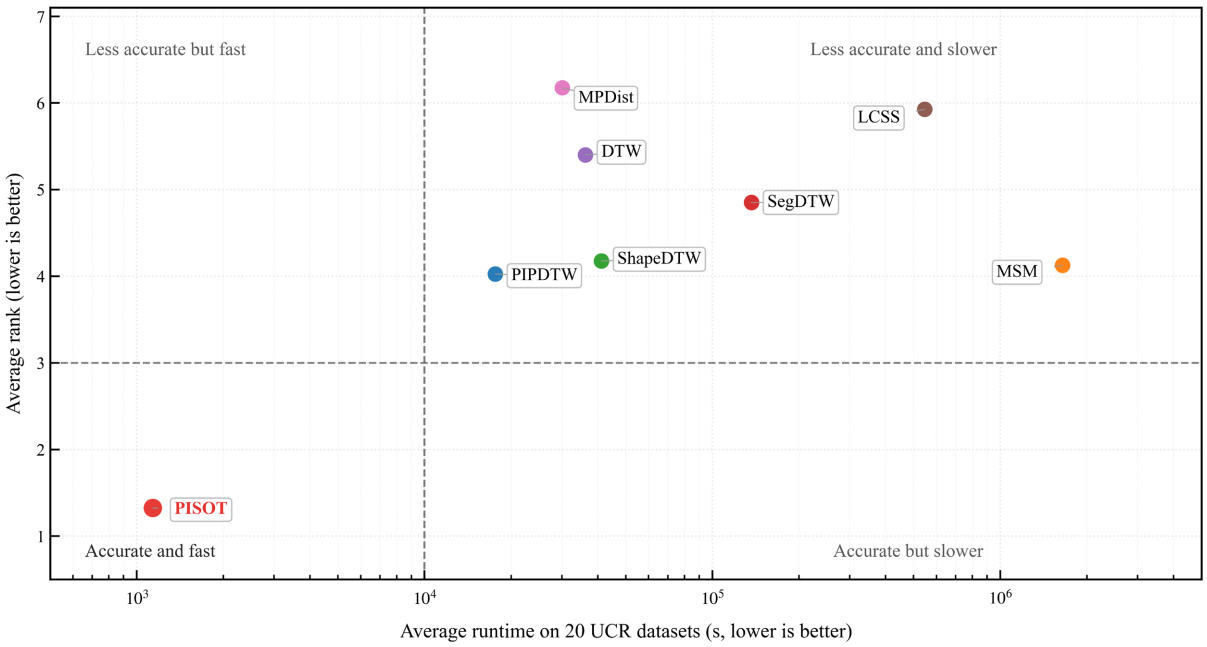

Figure 3. Comparison of average running time and accuracy ranking of 8 methods using the 1NN classifier
图 3. 1NN 分类器下的 8 种方法平均运行时间和准确率排名对比

由图 3 可以看出，PISOT 模型在计算效率与分类精度两个维度均展现出显著优势。PISOT 在 20 个 UCR 数据集上的平均运行时间最短，同时平均准确率排名最高。PISOT 计算效率相比传统 DTW 提速约 26 倍，相比最新子序列相似性度量方法 PIPDTW 提速约 13 倍。与此同时，PISOT 在分类准确率方面同样表现卓越，其平均排名显著优于所有对比方法。这充分验证了 PISOT 对于长序列数据分类具有实用性。

比较表 2 与表 3 可以发现 1NN 分类器下的 PISOT 在数据集上的准确率普遍高于质心分类器，但比

较图 2 与图 3 可以发现 1NN 分类器的计算时间明显长于质心分类器。这体现了分类任务中准确率与计算效率的权衡关系，质心分类器通过大幅减少计算次数显著提升计算效率，但可能因数据信息丢失而降低准确率，而 1NN 分类器虽计算成本较高，却能更充分地利用每个训练样本的判别信息，从而获得更高的分类准确率。

4.2.2. 统计检验

为进一步检验不同分类方法的性能差异，采用 Friedman 检验与 Nemenyi [36]事后检验，对 8 种方法在质心分类器与 1NN 分类器下的分类结果进行统计显著性分析。在 5%显著性水平下，对于 $d=8$ 种算法和 $\mathrm{K}=20$ 个数据集，Friedman 检验的临界值为 $F\left(d-1,(d-1)*(\mathrm{K}-1)\right)=F\left(7,133\right)=2.079$。在质心分类器下，计算得到统计量 $F_F=41.042$。在 1NN 分类器下，计算统计量 $F_F=52.350$。由于两种分类器下的统计量均明显大于临界值 2.067，说明 8 种方法之间的分类性能差异均达到了显著水平，即不同方法在两个分类框架下的表现并非随机波动所致，而是存在统计意义上的显著差异。

随后，采用 Nemenyi 事后检验来评估方法之间的两两差异。临界差 $CD=q_\alpha\sqrt{\mathrm{K}(\mathrm{K}+1)/6d}=2.334$。由图 4 和图 5 临界差异图可看出，在两种分类器下，PISOT 与所有 7 种对比方法均无连线，即两两差异均超过临界值，证明 PISOT 在统计意义上显著优于所有对比方法。这充分验证了本文提出的 PISOT 模型在不同分类框架下的性能优势，展现了其在多样化序列数据分类任务中的鲁棒性与优越性。

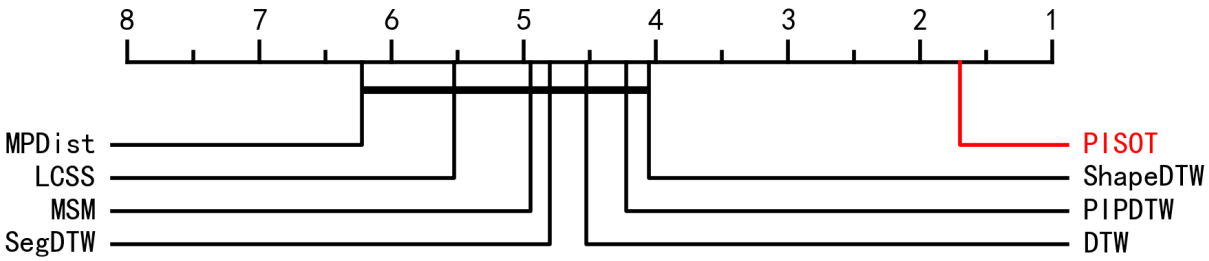

Figure 4. Critical difference diagram of eight methods based on the centroid classifier
图 4. 基于质心分类器 8 种方法临界差异图

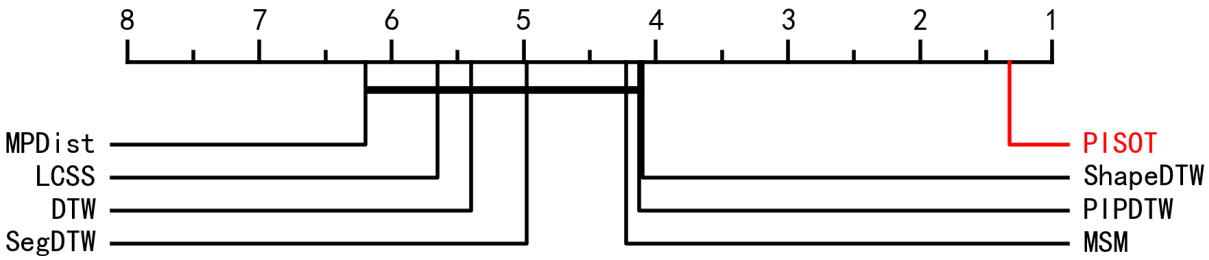

Figure 5. Critical difference diagram of eight methods based on the 1NN classifier
图 5. 基于 1NN 分类器 8 种方法临界差异图

4.2.3. 消融实验

为进一步评估模型的不同部分对分类准确率的影响，本文选取 10 个代表性的 UCR 数据集上进行了消融实验。在实验中，D_{seg} 代表滑动距离构建的代价矩阵，D_{CI_seg} 代表引入复杂度修正的滑动距离构建的代价矩阵，D_{pos} 代表位置惩罚正则项。本文设计了 4 组实验对照，分别评估 D_{seg}、加入复杂度修正因子的 D_{CI_seg}、加入位置惩罚正则项的 $D_{seg}+D_{pos}$ 以及同时加入复杂度修正因子和位置惩罚正则项的 $D_{CI\_seg}+D_{pos}$ 的分类准确率。

Table 4. Ablation study accuracy (%)
表 4. 消融实验的准确率(%)

数据编号	D_{seg}	D_{CI_seg}	$D_{seg}+D_{pos}$	$D_{CI\_seg}+D_{pos}$
ECG2	86.00	90.00	89.00	**91.00**
EOGVS	35.36	42.54	44.48	**48.62**
HE	45.31	51.56	53.12	**56.25**
HT	86.55	93.28	90.76	**95.80**
IS	43.45	48.55	44.55	**49.64**
Rock	76.00	78.00	80.00	**82.00**
SL	84.16	91.04	85.76	**91.52**
WS	49.53	73.20	51.72	**75.08**
AWR	90.67	95.33	93.67	**97.00**
HMD	18.92	28.38	22.97	**32.43**

由表 4 可见，完整模型 $D_{CI\_seg}+D_{pos}$ 在 10 个数据集上均取得最优分类准确率。具体分析表明，带复杂度修正因子的 D_{CI_seg} 相较于 D_{seg} 在 HT、HE 及 WS 等数据集上准确率有明显提升，说明复杂度修正因子能有效增强子序列的判别能力。同时，加入位置惩罚正则项 D_{pos} 后，在 EOGVS、HT 等数据集上模型性能进一步提升，表明位置惩罚正则项有效抑制子序列的错误对齐。综上，消融实验验证了本文所提出 D_{CI_seg} 与 D_{pos} 具有良好的协同优化效应，共同提升了模型对长序列数据分类的准确率。

4.2.4. 参数敏感性分析

为评估 PISOT 模型序列整体感知重要点个数 k 以及正则项系数 λ_{pos} 和 ε 在 1NN 分类器下对分类性能的影响，选取 ECG200 数据集进行敏感性分析。实验采用网格搜索与交叉验证法，逐一探究各参数对分类准确率的影响。

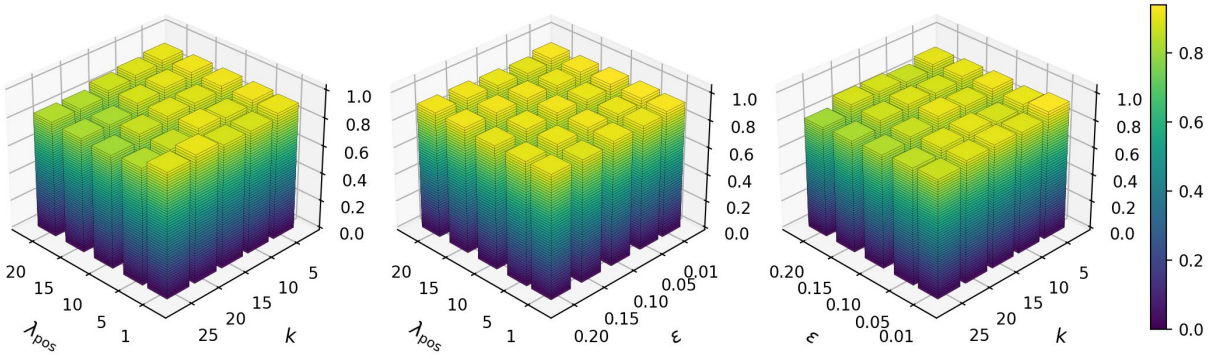

Figure 6. Hyperparameter sensitivity analysis on the ECG200 dataset
图 6. ECG200 数据集的超参数敏感性分析

如图 6 所示，当固定 k 时，随着 ε 值和 λ_{pos} 值增大，分类准确率逐渐下降。当固定 λ_{pos} 时，随 ε 的取值增大准确率下降，k 值对性能影响不大。当固定 ε 时，k 和 λ_{pos} 对分类准确率的影响不大。基于上述分析，将 ECG200 数据集参数 k 、λ_{pos} 和 ε 分别设为 10、1 和 0.01，以提升模型整体稳定性。

5. 结论与展望

本文提出一种面向长序列分类的变长子序列最优传输模型(PISOT)，用于解决长序列数据分类中存在的病态对齐、计算复杂度高及过度放大相似性等问题。该模型首先通过将 PIPs 与最优传输框架结合实现“多对多”匹配，缓解病态对齐的问题。其次，构建自适应变长子序列表示，将传统的逐点匹配扩展为子序列间的全局匹配，显著降低计算复杂度。最后，设计位置惩罚正则项，抑制位置过远的不合理匹配，从而缓解过度最大化相似性问题。在 UCR 公开数据集上的实验结果表明，PISOT 在长序列及高复杂度数据集上的分类准确率显著优于 LCSS、MSM、MPDist 等主流方法，计算效率也具有明显优势，统计显著性检验进一步验证了模型性能的稳定性和可靠性。

此外，PISOT 仍存在一定局限性。首先，模型涉及感知重要点个数、位置惩罚系数及熵正则化系数等多个参数，参数选择对模型性能存在一定影响，参数调优过程具有一定经验性和计算开销。其次，在短序列上的性能提升相对有限，由于短序列本身包含的时序信息较少，其优势尚不如长序列场景明显。未来可探索将最优传输模型计算的相似性度量作为深度学习模型的损失函数或正则化项，结合深度学习的强大表示能力与最优传输的优势，构建更加高效、鲁棒的序列数据分类模型，应用到工业设备多传感器故障诊断、语音识别以及人体动作识别等真实场景。

基金项目

河北省中央引导地方科技专项项目(246Z1825G)。

参考文献

[1] Toth-Laufer, E. and Batyrshin, I.Z. (2022) Similarity-Based Personalized Risk Calculation. 2022 *IEEE* 16*th International Symposium on Applied Computational Intelligence and Informatics* (*SACI*), Timisoara, 25-28 May 2022, 129-134. https://doi.org/10.1109/saci55618.2022.9919457

[2] Annapurna, N. and Sireesha, V. (2024) Optimizing Investment Selection through Similarity Measurement with Type-2 Intuitionistic Fuzzy Sets. *SN Computer Science*, **5**, Article No. 939. https://doi.org/10.1007/s42979-024-03285-3

[3] Gao, H., Huo, X., Jiang, Y., Zhu, C. and He, C. (2025) A DTW-Gaussian Spatiotemporal Self-Attention Network and Its Application in Industrial Fault Diagnosis with Unequal-Length Sensor Data. *IEEE Transactions on Industrial Informatics*, **21**, 6188-6197. https://doi.org/10.1109/tii.2025.3563521

[4] Sakoe, H. and Chiba, S. (1978) Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics*, *Speech*, *and Signal Processing*, **26**, 43-49. https://doi.org/10.1109/tassp.1978.1163055

[5] 魏国强, 周从华, 张婷. 基于多维分段和动态权重DTW的多元时间序列相似性度量方法[J]. 计算机与数字工程, 2021, 49(11): 2299-2304, 2406.

[6] Zhang, Q., Zhang, C., Cui, L., Han, X., Jin, Y., Xiang, G., *et al*. (2022) A Method for Measuring Similarity of Time Series Based on Series Decomposition and Dynamic Time Warping. *Applied Intelligence*, **53**, 6448-6463. https://doi.org/10.1007/s10489-022-03716-9

[7] 张艳, 史晨辉, 苏美红, 等. 基于邻域形状的时间序列相似性度量[J]. 计算机工程与设计, 2025, 46(6): 1578-1585.

[8] Li, H. (2021) Time Works Well: Dynamic Time Warping Based on Time Weighting for Time Series Data Mining. *Information Sciences*, **547**, 592-608. https://doi.org/10.1016/j.ins.2020.08.089

[9] Hao, S., Wang, Z. and Yuan, J. (2023) Local Morphological Patterns for Time Series Classification. *Intelligent Data Analysis*, **27**, 653-674. https://doi.org/10.3233/ida-216548

[10] Hong, J.Y., Park, S.H. and Baek, J. (2020) SSDTW: Shape Segment Dynamic Time Warping. *Expert Systems with Applications*, **150**, Article ID: 113291. https://doi.org/10.1016/j.eswa.2020.113291

[11] Li, J., Zhao, J., Wang, F., Kawata, S., Chugo, D. and She, J. (2025) Features Based Dynamic Time Warping of Multidimensional Series for Aligning Sensor Data. *IECON* 2025—51*st Annual Conference of the IEEE Industrial Electronics Society*, Madrid, 14-17 October 2025, 1-6. https://doi.org/10.1109/iecon58223.2025.11221538

[12] Qiu, L., Qiu, C. and Song, C. (2024) ESDTW: Extrema-Based Shape Dynamic Time Warping. *Expert Systems with Applications*, **239**, Article ID: 122432. https://doi.org/10.1016/j.eswa.2023.122432

[13] 金俊超, 马昌忠, 陈国良, 等. 基于 UCR-DTW 的地磁序列定位算法[J]. 合肥工业大学学报(自然科学版), 2021,

44(11): 1551-1556.

[14] Lahreche, A. and Boucheham, B. (2021) A Fast and Accurate Similarity Measure for Long Time Series Classification Based on Local Extrema and Dynamic Time Warping. *Expert Systems with Applications*, **168**, Article ID: 114374. https://doi.org/10.1016/j.eswa.2020.114374

[15] Ge, L. and Chen, S. (2020) Exact Dynamic Time Warping Calculation for Weak Sparse Time Series. *Applied Soft Computing*, **96**, Article ID: 106631. https://doi.org/10.1016/j.asoc.2020.106631

[16] Chaerul Haviana, S.F. (2015) Sistem Gesture Accelerometer Dengan Metode Fast Dynamic Time Warping (FastDTW) *Jurnal Sistem Informasi Bisnis*, **5**, 151-160. https://doi.org/10.21456/vol5iss2pp151-160

[17] Tan, C.W., Petitjean, F. and Webb, G.I. (2019) Elastic Bands across the Path: A New Framework and Method to Lower Bound DTW. In: Berger-Wolf, T. and Chawla, N., Eds., *Proceedings of the* 2019 *SIAM International Conference on Data Mining*, *Society for Industrial and Applied Mathematics*, Alberta, 2-4 May 2019, 522-530. https://doi.org/10.1137/1.9781611975673.59

[18] Webb, G.I. and Petitjean, F. (2021) Tight Lower Bounds for Dynamic Time Warping. *Pattern Recognition*, **115**, Article ID: 107895. https://doi.org/10.1016/j.patcog.2021.107895

[19] Ye, L. and Keogh, E. (2009) Time Series Shapelets: A New Primitive for Data Mining. *Proceedings of the* 15*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, 28 June-July 2009, 947-956. https://doi.org/10.1145/1557019.1557122

[20] Gharghabi, S., Imani, S., Bagnall, A., Darvishzadeh, A. and Keogh, E. (2018) Matrix Profile XII: MPdist: A Novel Time Series Distance Measure to Allow Data Mining in More Challenging Scenarios. 2018 *IEEE International Conference on Data Mining* (*ICDM*), Singapore, 17-20 November 2018, 965-970. https://doi.org/10.1109/icdm.2018.00119

[21] Hirschberg, D.S. (1977) Algorithms for the Longest Common Subsequence Problem. *Journal of the ACM*, **24**, 664-675. https://doi.org/10.1145/322033.322044

[22] Stefan, A., Athitsos, V. and Das, G. (2013) The Move-Split-Merge Metric for Time Series. *IEEE Transactions on Knowledge and Data Engineering*, **25**, 1425-1438. https://doi.org/10.1109/tkde.2012.88

[23] Zhao, J. and Itti, L. (2018) ShapeDTW: Shape Dynamic Time Warping. *Pattern Recognition*, **74**, 171-184. https://doi.org/10.1016/j.patcog.2017.09.020

[24] Mishra, K., Basu, S. and Maulik, U. (2021) SeqDTW: A Segmentation Based Distance Measure for Time Series Data. *Transactions of the Indian National Academy of Engineering*, **6**, 709-730. https://doi.org/10.1007/s41403-021-00230-1

[25] 郝石磊, 王志海, 刘海洋. 基于局部梯度和二进制模式的时间序列分类算法[J]. 软件学报, 2022, 33(5): 1817-1832.

[26] 胡萌窈, 王鹏, 汪卫. 基于子序列相似性的时间序列在线状态识别[J/OL]. 计算机应用与软件: 1-8. https://link.cnki.net/urlid/31.1260.TP.20240725.1401.007, 2026-01-22.

[27] Zhang, H., Feng, J., Li, J. and Yao, Q. (2024) Efficient Top-*k* DTW-Based Sensor Data Similarity Search Using Perceptually Important Points and Dual-Bound Filtering. *IEEE Sensors Journal*, **24**, 41231-41242. https://doi.org/10.1109/jsen.2024.3478214

[28] Villani, C. (2021) Topics in Optimal Transportation. Rhode Island: American Mathematical Society.

[29] Zhang, Z., Tang, P. and Corpetti, T. (2020) Time Adaptive Optimal Transport: A Framework of Time Series Similarity Measure. *IEEE Access*, **8**, 149764-149774. https://doi.org/10.1109/access.2020.3016529

[30] Su, B. and Hua, G. (2019) Order-preserving Optimal Transport for Distances between Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**, 2961-2974. https://doi.org/10.1109/tpami.2018.2870154

[31] Latorre, F., Liu, C., Sahoo, D. and Hoi, S.C.H. (2023) OTW: Optimal Transport Warping for Time Series. *ICASSP* 2023—2023 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*), Rhodes Island, 4-10 June 2023, 1-5. https://doi.org/10.1109/icassp49357.2023.10095915

[32] Chung, F., Fu, T., Luk, R.W. and Ng, V.T.Y. (2001) Flexible Time Series Pattern Matching Based on Perceptually Important Points. *Workshop on Learning from Temporal and Spatial Data in International Joint Conference on Artificial Intelligence*, Seattle, 6-12 August 2001, 1-7.

[33] Batista, G.E.A.P.A., Keogh, E.J., Tataw, O.M. and de Souza, V.M.A. (2013) CID: An Efficient Complexity-Invariant Distance for Time Series. *Data Mining and Knowledge Discovery*, **28**, 634-669. https://doi.org/10.1007/s10618-013-0312-3

[34] Dau, H.A., Bagnall, A., Kamgar, K., Yeh, C., Zhu, Y., Gharghabi, S., Ratanamahatana, C.A. and Keogh, E.J. (2018) The UCR Time Series Archive. *IEEE*/*CAA Journal of Automatica Sinica*, **6**, 1293-1305.

[35] 孟晓静, 万源. 自适应代价动态时间弯曲的多元时间序列相似性度量[J]. 统计与决策, 2020, 36(2): 25-29.

[36] Demiar, J. and Schuurmans, D. (2006) Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, **7**, 1-30.