

# Regularization-free Diffeomorphic Temporal Alignment Nets



# 目录 CONTENTS

---

一、摘要

二、引言

三、

四、

五、

六、

## 一、摘要

## 1、摘要

在时间序列分析中，非线性时间错位是一个主要问题，甚至阻碍了简单的平均。

针对此问题的一种有效的基于学习的解决方案是微分同胚时间对齐网络 (DTAN)，它通过依赖微分同胚时间变换网络和联合对齐任务的摊销，消除了缺点传统的对齐方法。

现有的 DTAN 公式很大程度上依赖于正则化项，其最佳超参数是特定于数据集的，并且通常通过大量实验进行搜索。

本文提出了一种无正则化的 DTAN，它消除了执行如此昂贵且通常不切实际的搜索的需要。具体来说，提出了一种新的表现良好的损失，称为逆一致性平均误差 (ICAE)，以及相关的新三元组损失。

对 128 个 UCR 数据集进行的广泛实验表明，尽管未使用正则化，所提出的方法仍优于当代方法。此外，ICAE 还产生了第一个支持可变长度信号的 DTAN。

代码可在 <https://github.com/BGU-CS-VIL/RF-DTAN> 获取。

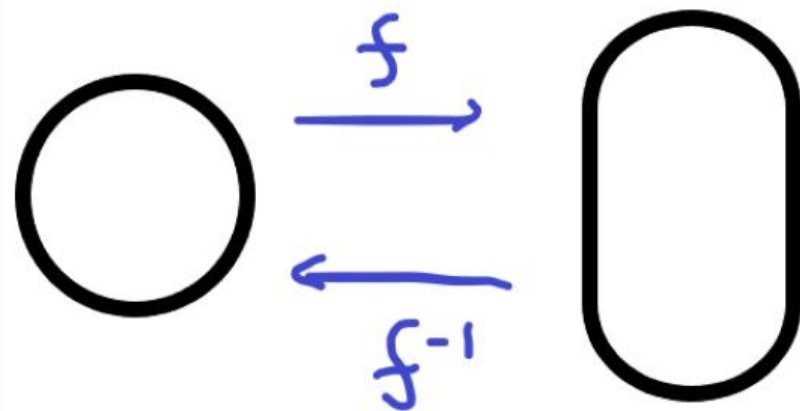
## 二、引言

## 1、微分同胚

同胚就是一个拥有连续逆映射的连续映射，就像一个圆圈和一个正方形那样；

在同胚的基础上，如果我们进一步要求我们的同胚映射可微，且其逆映射可微，我们就得到了微分同胚的概念，就像一个圆和一个椭圆。

微分同胚本身的重要性在于，在它把两个光滑形状上的点等同起来的同时，它也把对应的点处其他的光滑结构等同起来了。微分同胚相当一部分的重要意义在于，你可以把一个空间“当成”另一个空间来处理，尽管在处理长度，面积这些概念时，你也许需要在局部乘以适当的系数；例如当你在研究球面时，你其实可以去研究球面上某些区域对应的参数平面，而后者明显要轻松很多。



$f$  是连续的。  
 $f^{-1}$  是连续的。

$f, f^{-1}$  是光滑的。（无穷可微）

## 2、DTW

$Q = q_1, q_2, \dots, q_i, \dots, q_n$ ;

$C = c_1, c_2, \dots, c_j, \dots, c_m$ ;

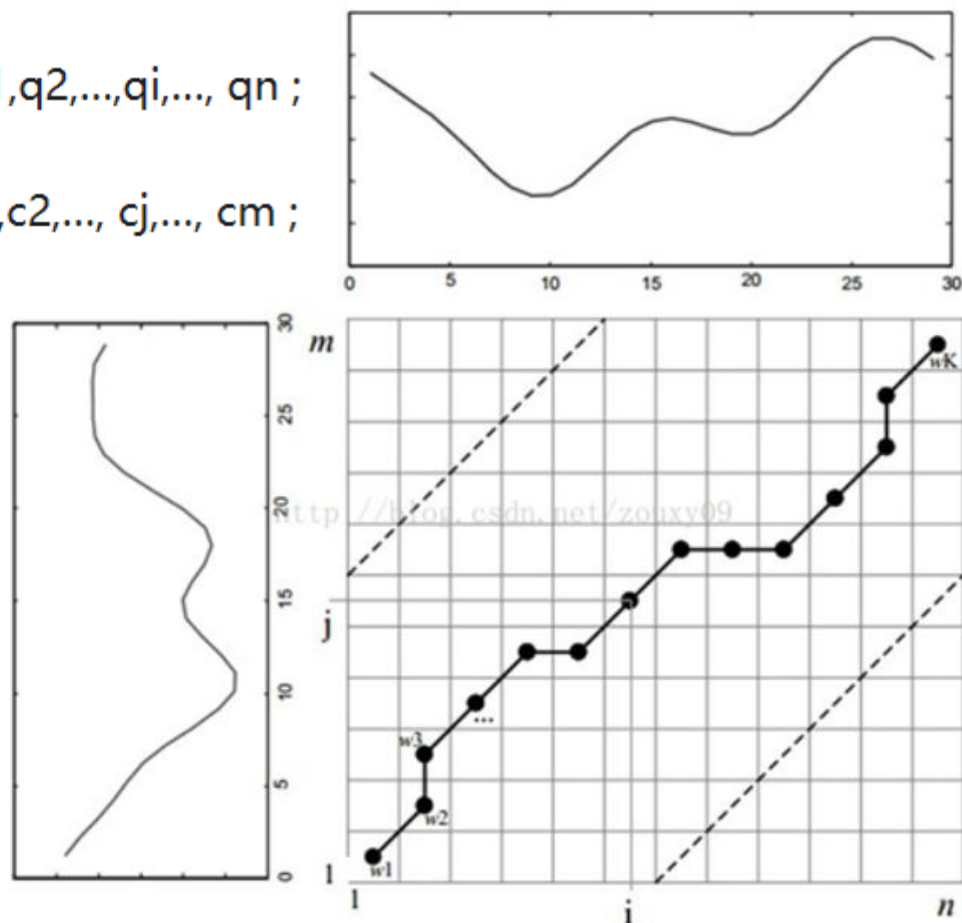
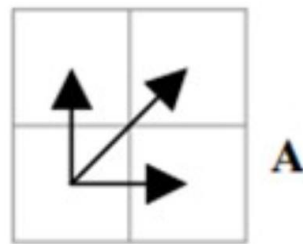


Figure 3: An example warping path.

动态时间规整DTW是一个典型的优化问题，它用满足一定条件的的时间规整函数 $W(n)$ 描述测试模板和参考模板的时间对应关系，求解两模板匹配时累计距离最小所对应的规整函数。

首先，采用两个序列中每一对“点”之间的距离来计算形似度，即使两个序列中的点的个数可能不一样。不过，因为可以warping规整时间轴，所以，我们并不是在两个序列中依次取一对点来计算距离，而是每个点有可能对应于另一个序列中的多个点。



### 3、问题和解决

由于微分同胚的空间很大，并且为了阻止不利的解决方案，通常会在目标函数中添加一些正则化项

$$T^* = \arg \min_{T \in \mathcal{T}} D(u_i, u_j \circ T) + \mathcal{R}(T; \lambda)$$

在  $N$  个信号集合的情况下，成对方法通常不能很好地推广，容易出现漂移错误，并且可能会引入不一致的解决方案。这激发了联合对齐 (JA) 方法

$$(T_i^*)_{i=1}^N, \mu = \arg \min_{(T_i)_{i=1}^N \in \mathcal{T}, u} \sum_{i=1}^N D(u, u_i \circ T_i) + \mathcal{R}(T_i; \lambda)$$

$\mu$  是潜在信号，概念上被认为是系综的平均信号（或质心）

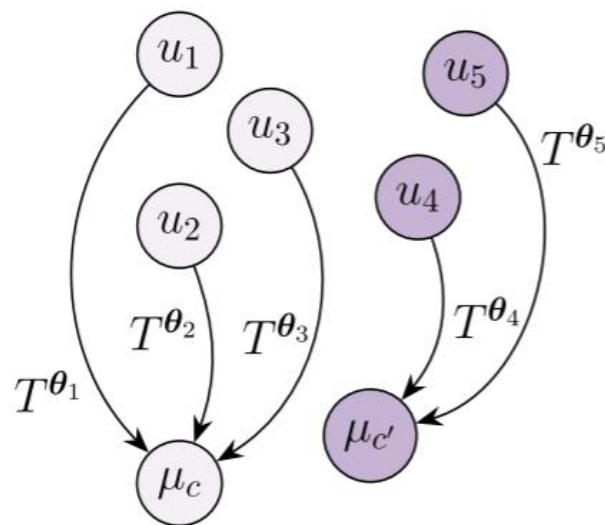
JA 方法（包括深度学习 (DL) 方法）的成功关键取决于  $\mathcal{R}(\cdot; \lambda)$  的选择，更重要的是其 HP（超参数）、 $\lambda$  的选择。



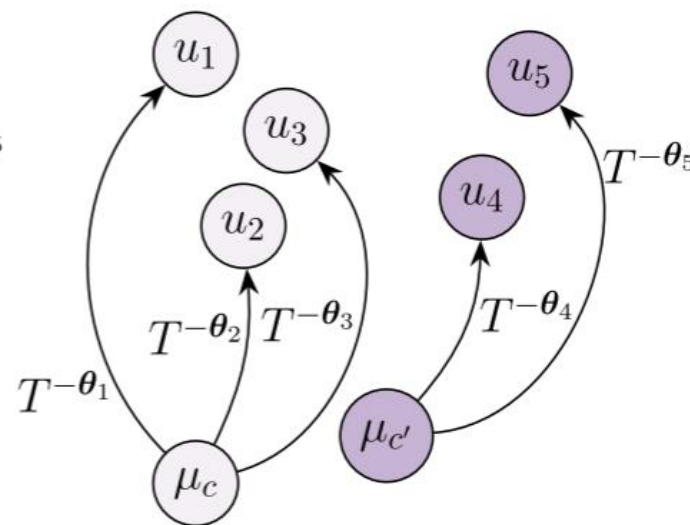
### 3、问题和解决

提出了一种基于新损失的无正则化深度学习方法，称为逆一致性平均误差（ICAE），用于时间序列 JA 和平均。

减轻了对扭曲正则化的需求，并且可以在任何 JA 方法中使用，只要扭曲是可逆的。



(a) Centroids computed using forward warps



(b) The ICAE loss computed using backward warps

## 4、贡献

- 1) 引入 JA 和平均任务的 ICAE 损失，从而消除对预测扭曲使用正则化的需要。
- 2) 所提出的损失的三元组损失变体，以实现更好的类间分离。
- 3) JA 的显式公式和可变长度数据的平均。
- 4) 在 UCR 时间序列分类档案中的 128 个数据集上设置新的最先进 (SOTA) 结果。

### 三、方法

## 1、现有方法

提出了一种针对时间序列 JA 和使用 DTAN 进行平均的无正则化方法。利用了  $\tau$  是微分同胚族，因此元素是可逆的。

领先的 JA 方法依赖于扭曲正则化来避免不切实际的变形和/或琐碎的解决方案。然而，其最佳 HP 是特定于数据集的。

由于时间序列数据在不同的应用领域之间差异很大（例如，心电图与音频记录相比），因此很难确定合适的  $\lambda$  值。

## 2、时间变压器网络

TTN 预测扭曲参数  $\theta$  并将  $T^\theta$  应用于输入信号，由三个模块组成。

第一个是所谓的本地化网络。这是一个神经网络，用  $\text{floc}(\cdot)$  表示，它以一批序列  $(u_i)_{i=1}^N$  作为输入，并预测相应的扭曲参数  $(\theta_i)_{i=1}^N$ 。

第二个是网格生成器，它创建一个由均匀间隔的点组成的网格  $G \subset \Omega$ ，然后由  $T^{\theta_i}$  扭曲。最后，网格采样器通过使用  $u_i$  在  $(T^{\theta_i})^{-1}(G)$  处插值其值来计算扭曲信号  $v_i = u_i \circ T^{\theta_i}$ 。

在这项工作中，将  $\text{floc}$  设置为 InceptionTime，而不是使用的时间卷积网络 (TCN)。

InceptionTime 最初是为时间序列分类而设计的，它受到 Inception-v4 架构的启发，由多个利用图像分类中流行的瓶颈设计的 Inception 模块组成。

我们在  $\text{floc}$  的倒数第二层之前合并了全局平均池 (GAP) 算子，这使得模型的可训练参数数量 w.r.t. 保持固定和输入大小保持固定。

### 3、逆一致性平均误差

逆一致性误差 (ICE) 给定两个图像  $I_1$  和  $I_2$ ，分别具有域  $\Omega_1$  和  $\Omega_2$ ，潜在空间映射  $f_1: \Omega_2 \rightarrow \Omega_1$  和  $f_2: \Omega_1 \rightarrow \Omega_2$  应该是一致的；即， $f_2 = f_1^{-1}$  且  $f_1 = f_2^{-1}$ 。

$$\int_{\Omega_2} \|f_2(f_1(x)) - x\|^2 dx + \int_{\Omega_1} \|f_1(f_2(x)) - x\|^2 dx$$

提出了一种新形式的逆一致性，使其对于 JA 任务也很有用。与原始 ICE 不同的是，原始 ICE 是成对的，充当添加到主要损失中的正则化项，LICAE 测量估计平均序列与其各自组成员之间的一致性。此外，LICAE 本身就是整个损失，而不是添加到另一个损失中的正则化项。也就是说，(ICE 的) 泛化本身就是一个专用损失函数，并导致一致的 JA。重要的是，它消除了使用任何形式的正则化的需要，而这反过来又消除了（微不足道的）调整正则化 HP 的需要。

### 3、逆一致性平均误差

JA 损失函数 对于单个类别，损失是对齐信号的方差：

$$\mathcal{L}_{\text{data}} \triangleq \frac{1}{N} \sum_{i=1}^N \left\| u_i \circ T^{\theta_i} - \mu \right\|_{\ell_2}^2$$

其中  $\| \cdot \|_{\ell_2}$  是  $\ell_2$  范数， $(\theta_i)_{i=1}^N$  是 floc 预测的扭曲参数。

$$\mu = \frac{1}{N} \sum_{i=1}^N u_i \circ T^{\theta_i}$$

是对齐后的平均信号。

在多元情况下，损失是类内方差之和，通常称为类内平方和 (WCSS)：

$$\mathcal{L}_{\text{data}} \triangleq \sum_{k=1}^K \frac{1}{N_K} \sum_{i:y_i=k} \left\| u_i \circ T^{\theta_i} - \mu_k \right\|_{\ell_2}^2$$
$$\mu_k = \frac{1}{N_K} \sum_{i:y_i=k} u_i \circ T^{\theta_i}$$

是  $k$  类的对齐后平均值。

### 3、逆一致性平均误差

很明显，在存在这些损失的情况下，需要扭曲正则化项  $R(T\theta_i; \lambda)$ 。首先， $L_{data}$  并不鼓励扭曲一致性。其次，可以通过严重扭曲信号来减少方差（甚至为零！），并且这个问题只会由于插值伪影而变得更糟。然而，最佳正则化是特定于数据集的。例如，在许多情况下，惩罚太大的变形可能并不理想。同样，在时间平滑度先验的情况下，很难确定“正确”的平滑度。

正则化对使用 DBA、SoftDTW 和 DTAN 进行重心计算的关键作用。 $\gamma$ （对于 SoftDTW）或  $\lambda\Sigma$ 、 $\lambda smooth$ （对于 DTAN）的不正确值通常会导致不切实际的扭曲或过度限制扭曲（例如，DTAN 的强先验）。



### 3、逆一致性平均误差

提出了一种新的损失，当平均序列既是方差的最小值又与其类别一致时，该损失就会最小化。逆一致性平均误差损失（ICAEL），定义为：

$$\mathcal{L}_{\text{ICAEL}} \triangleq \sum_{k=1}^K \frac{1}{N_K} \sum_{i:y_i=k} \left\| \mu_k \circ T^{-\theta_i} - u_i \right\|_{\ell_2}^2$$

ICAEL 使用反向扭曲  $T^{-\theta_i}$  来测量平均信号  $\mu_k$  与同类信号  $u_i$  的拟合程度。它首先使用预测的扭曲对齐  $k$  类中的所有信号，然后计算它们的平均值  $\mu_k$ ，最后使用  $T^{-\theta_i}$  将  $\mu_k$  扭曲回每个  $u_i$ ，从而确保它们之间的一致性。

一个关键的见解是，强烈建议不要采用琐碎的解决方案或不切实际的扭曲，因为这会导致  $\mu_k$  的估计值很差，进而导致它与原始信号之间产生很大的差异。换句话说，损失有利于现实变形，而不需要添加正则化项。

### 3、逆一致性平均误差

---

**Algorithm 1** The JA training with an ICAE loss

---

**Input:**  $N_{\text{epochs}}, f_{\text{loc}}$

**Data:**  $(u_i, y_i)_{i=1}^N$

**Output:**  $f_{\text{loc}}(\cdot)$ , trained for joint alignment

**for** *each epoch and each batch*  $j \in \{1, \dots, N_{\text{batches}}\}$  **do**

$\mathcal{L}_{\text{batch}} \leftarrow 0$

$(u_i, y_i)_{i=1}^{N_j} \leftarrow \text{batch}_j$

$(\theta_i)_{i=1}^{N_j} \leftarrow (f_{\text{loc}}(u_i))_{i=1}^{N_j}$

**for**  $k \in \{1, \dots, K\}$  **do**

$\mu_k = \frac{1}{N_k} \sum_{i:y_i=k} (u_i \circ T^{\theta_i})$

$\mathcal{L}_{\text{ICAE}} = \frac{1}{N_K} \sum_{i:y_i=k} \|\mu_k \circ T^{-\theta_i} - u_i\|_{\ell_2}^2$

$\mathcal{L}_{\text{batch}} \mathrel{+}= \mathcal{L}_{\text{ICAE}}$

Perform an optimization step to minimize  $\mathcal{L}_{\text{batch}}$

---

## 4、质心三重损失

虽然 LICAE 意味着一致性，但它对不同类之间的分离是不可知的。也就是说，虽然 DTW 等指标完全是数据驱动的，但基于学习的模型可以用来学习任务驱动的表达。将质心三重损失引入到框架中以鼓励类间分离。我们将三元组  $(u_i^a, \mu^p, \mu^n)$  上的逆一致质心三元组损失定义为：

$$\mathcal{L}_{\text{ICAE-triplet}}(u_i^a, \mu^p, \mu^n) \triangleq \max(0, \|\mu^p - \mu^n\|_{\ell_2} - \alpha)$$

其中  $\mu^p$ 、 $\mu^n$  分别是正类质心和负类质心， $\alpha$  是它们之间的间隔。由于  $\mu^p$  和  $\mu^n$  都是通过反向扭曲进行比较的，LICAE-三元组不会破坏样本与其平均值之间的一致性。LICAE-三联体与 LICAE 串联使用。

## 5、可变长度联合对齐

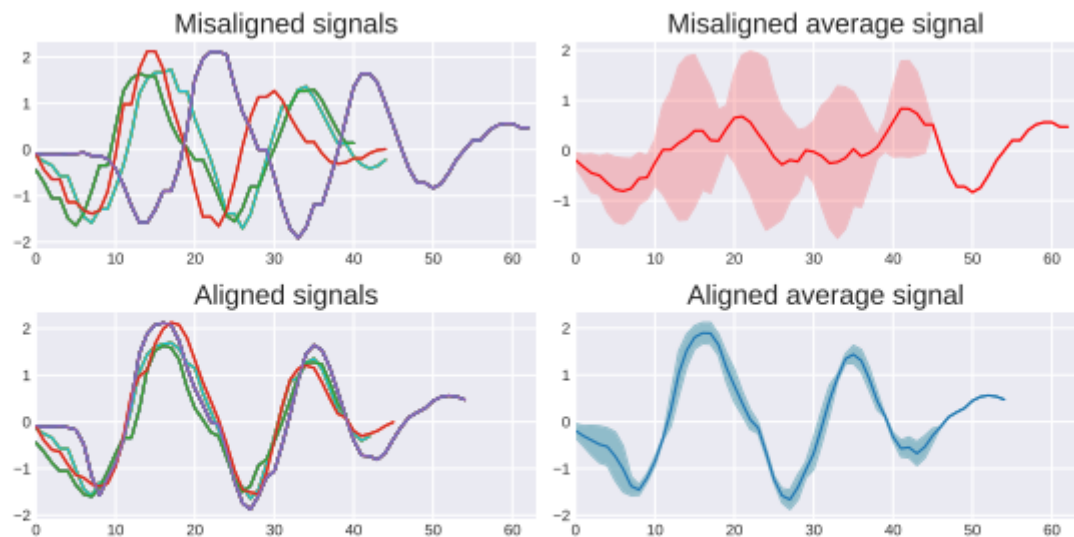


Figure 4. JA of variable-length data (Dataset: *ShakeGestureWiimoteZ*) using the proposed  $\mathcal{L}_{\text{ICAE}}$ . Shaded area is  $\pm$  std. dev.

提出的 LICAE 还允许 JA 和可变长度序列的平均，而无需使用专门的损失函数或调整  $T\theta$  上的边界条件。可以处理固定长度和可变长度的数据。它通过以下方式实现这一点。

首先，通过在每个时间步将相关值的总和除以非缺失值的数量来生成对齐后平均信号。也就是说，对于沿平均信号  $\mu$  的持续时间的每个时间步长  $t$ ，我们计算：

$$\mu[t] = \frac{1}{N_{\text{valid}}} \sum_{i: (u_i \circ T^{\theta_i})[t] \neq \text{null}} (u_i \circ T^{\theta_i})[t]$$

其中  $N_{\text{valid}}$  是其域包括映射到  $t$  的点的信号数量。然后，当  $\mu$  向后扭曲时，无需修改即可计算公式 12。从实现的角度来看，我们注意到输入和/或损失中的任何空值都会破坏计算图。为了避免 for-loops 并批量计算反向传播，首先用零填充所有样本（相对于最长信号）并为缺失值创建指示符掩码在计算上是有效的。掩模也会因公式 14 中的  $T\theta$  而变形。

## 1、结论

提出了逆一致性平均误差 (LICA E)，这是一种新颖的损失函数，用于通过微分同胚时间变换网络进行无正则化时间序列联合对齐和平均。

该方法利用微分同胚扭曲的可逆性并产生有效的 JA，同时减轻了广泛 HP 搜索的需要。我们还提出了 LICA E-三元组，它允许使用三元组质心损失的扭曲一致变体来实现更好的类间分离。

此外，我们通过所提出的框架引入了可变长度时间序列数据的联合对齐的公式。对 128 个数据集进行的广泛实验证明了我们方法的有效性，在不需要扭曲正则化的情况下实现了 SOTA 性能。最后，我们的方法还可以与另一种无正则化的联合对齐方法结合使用。