

# Shape and Time Distortion Loss for Training Deep Time Series Forecasting Models

Machine Learning For Time Series

---

Ben Kabongo & Martin Brosset

December 18, 2023

Master MVA, ENS Paris-Saclay

# Table of contents

1. Introduction

2. Method

3. Data

4. Results

5. Conclusion

# Introduction

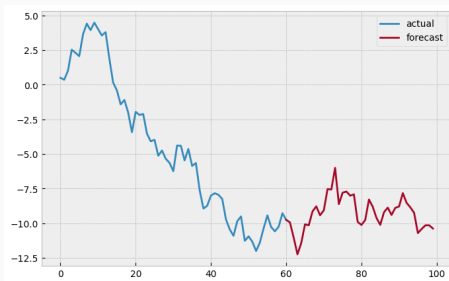
---

# Prévision des séries temporelles

Analyser la dynamique et les corrélations entre les données historiques afin de prédire le comportement futur.

Domaine de recherche très actif en machine/deep learning.

**Exemples de problématiques** : maîtrise des signaux non stationnaires, prédiction précise à plusieurs étapes, etc.



**Figure 1:** Times series forecasting

# MSE (Erreur quadratique moyenne)

Loss utilisée dans la majorité de modèles proposés pour la prévision des séries temporelles.

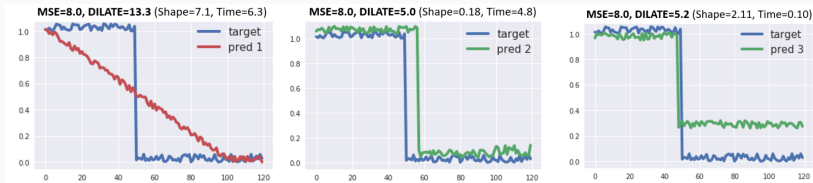


Figure 2: Limites de la MSE (tiré de [1])

## Limites :

- Capture des dynamiques complexes et des transitions abruptes propres aux séries temporelles
- Tend à lisser les prédictions, ce qui peut nuire à la détection des pics ou des changements soudains

# DTW (Dynamic Time Warping)

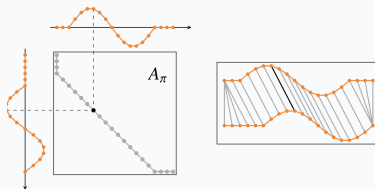
Alternative à la MSE.

## Avantages

- Tiens compte des distorsions possibles dans le temps.
- Capable de mesurer avec précision la similarité entre deux séquences temporelles.

## Limites:

- Ne considère pas les faibles décalages temporels.
- **N'est pas différentiable !**
- Difficile à intégrer dans les réseaux de neurones.



**Figure 3:** DTW

# DILATE (DIstortion Loss including shApe and TimE)

- **DILATE** : proposition d'une fonction de perte différentiable.
- Intègre à la fois la forme et la temporalité des séries temporelles.
- DILATE pallie les insuffisances de la MSE et de la DTW classique.

## Répartition des tâches

- **Ben :**

Trouver et étudier des datasets de séries temporelles aux propriétés différentes.

Adaptation d'une partie du code original.

- **Martin :**

Lancer le modèles sur les différents datasets en optimisant les paramètres et extrayant les résultats.

**Code:** 60% de code réutilisé.

<https://github.com/vincent-leguen/DILATE>



# Method

---

On souhaite prédire la suite des séries temporelles sur  $k$  pas de temps.

On note :

- $\{\mathbf{x}_i\}_{i=1,\dots,N}$  : ensemble d'apprentissage.
- $\mathbf{x}_i = (x_i^1, \dots, x_i^n) \in \mathbb{R}^{d \times n}$  : série temporelle de taille  $n$  et de dimension  $d$ .
- $\hat{\mathbf{y}}_i = (\hat{y}_i^1, \dots, \hat{y}_i^k)$  : prédiction pour la série temporelle  $\mathbf{x}_i$ .
- $\mathbf{y}_i^* = (y_i^{*1}, \dots, y_i^{*k})$  : suite réelle de la série temporelle  $\mathbf{x}_i$ .

# DILATE vue d'ensemble

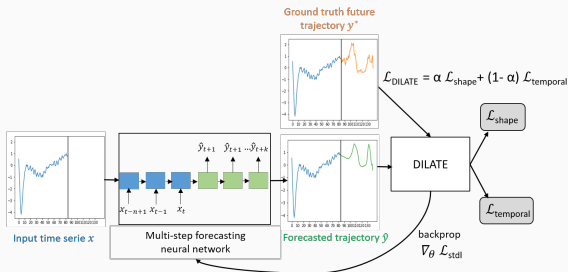


Figure 4: DILATE Loss

$(\hat{y}_i, y_i^*)$  : couple prédiction et suite réelle de la série temporelle  $x_i$ .

$$L_{dilate} = \alpha L_{shape} + (1 - \alpha) L_{temporal}$$

La loss DILATE se décompose en deux termes distincts : un terme de forme (**shape**) et un terme temporel (**temporal**)

## Terme de forme (Shape Loss)

Le terme de forme est noté  $L_{shape}$ .

Basé sur la Dynamic Time Warping (DTW).

$$DTW(\hat{y}_i, y_i^*) = \min_{A \in \mathcal{A}_{k,k}} \langle A, \Delta(\hat{y}_i, y_i^*) \rangle$$

Rendu différentiable grâce à l'introduction d'un opérateur de minimum lisse  $\min_{\gamma}$  ( $\gamma > 0$ ) :

$$\min_{\gamma}(a_1, \dots, a_n) = -\gamma \log \left( \sum_{i=1}^n \exp \left( -\frac{a_i}{\gamma} \right) \right)$$

Le terme de forme différentiable est donné par :

$$L_{shape}(\hat{y}_i, y_i^*) = DTW_{\gamma}(\hat{y}_i, y_i^*) := -\gamma \log \left( \sum_{A \in \mathcal{A}_{k,k}} \exp \left( -\frac{\langle A, \Delta(\hat{y}_i, y_i^*) \rangle}{\gamma} \right) \right)$$

# Terme temporel (Temporal Loss)

Le terme de forme est noté  $L_{temporal}$ .

Pénalise les distorsions temporelles entre les séries prédites et les séries réelles.

Inspiré de l'Indice de Distorsion Temporelle (TDI). Utilise une matrice  $\Omega$  pour quantifier les écarts temporels.

$$\Omega(h, j) = \frac{1}{k^2}(h - j)^2$$

Le terme temporel différentiable est donné par :

$$L_{temporal}(\hat{y}_i, y_i^*) := \frac{1}{Z} \sum_{A \in \mathcal{A}_{k,k}} \langle \Omega, A \rangle \exp \left( -\frac{\langle A, \Delta(\hat{y}_i, y_i^*) \rangle}{\gamma} \right)$$

Avec :

$$Z = \sum_{A \in \mathcal{A}_{k,k}} \exp \left( -\frac{\langle A, \Delta(\hat{y}_i, y_i^*) \rangle}{\gamma} \right)$$

## Forward et Backward pass

Le calcul de  $\mathcal{A}_{k,k}$  pose un problème car est très long à calculer, car croissant exponentiellement avec  $k$ .

**Forward pass**  $L_{shape}$  : Avec la méthode de programmation dynamique de Bellman, trouver le chemin optimal de distorsion minimale dans  $\mathcal{A}_{k,k}$  sans avoir à évaluer tous les chemins possibles.

**Backward pass**  $L_{shape}$  : Lors du passage arrière, l'implémentation personnalisée dans Pytorch prend avantage des résultats intermédiaires obtenus pendant le passage avant.

**Forward pass**  $L_{temporal}$  : Pour le calcul du gradient de la soft DTW, on utilise la même implémentation que le passage arrière pour  $L_{shape}$ .

**Backward pass**  $L_{temporal}$  : Calcul du Hessien de la soft DTW avec une méthode de programmation dynamique, rendant le processus plus rapide que l'auto-différenciation standard.

Soit  $\alpha \in [0, 1]$ , la fonction de perte DILATE est donnée par :

$$L_{dilate} = \alpha L_{shape} + (1 - \alpha) L_{temporal} \\ := -\gamma \log \left( \sum_{A \in \mathcal{A}_{k,k}} \exp \left( -\frac{\langle A, \alpha \Delta(\hat{y}_i, y_i^*) + (1 - \alpha) \Omega \rangle}{\gamma} \right) \right)$$

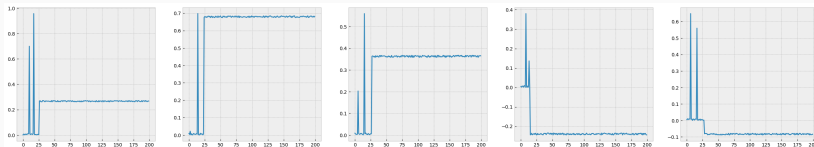
L'intérêt de cette fonction de loss par rapport aux travaux qui ont précédés l'article qui implémentent une DTW différentiable est que l'ajout de la pénalisation de distortion temporelle.

# Data

---



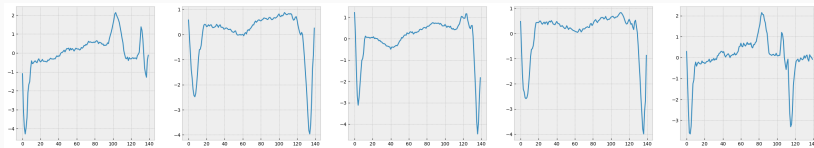
# Données synthétiques



**Figure 5:** Données synthétiques

Les données synthétiques sont utilisées pour évaluer les différentes loss sur les changements soudains dans les séries temporelles.

Les signaux comprennent donc deux pics placés aléatoirement pour chaque signal afin d'incorporer les changements brusques.



**Figure 6:** Dataset ECG 5000

Obtenues à partir de la base de données BIDMC Congestive Heart Failure Database (CHFDB) de PhysioNet.

Prétraitées en deux temps : chaque battement de cœur a d'abord été extrait, puis interpolé pour avoir la même longueur de 140 pas de temps.

5 classes de battements cardiaques différents dans le dataset : normal (1), et d'autres classements anormaux repartis en quatre classes différentes.

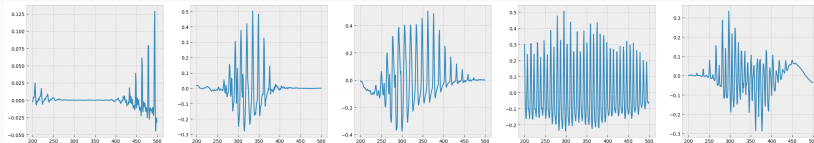


**Figure 7:** Dataset Traffic

Comprend les taux d'occupations des routes, normalisé entre 0 et 1, du département des transports de Californie, sur 48 mois à partir de 2015-2016.

Nous sélectionnons 1000 séries temporelles d'une longueur de 192 : 500 pour l'apprentissage et 500 pour l'évaluation.

# Insect Sound



**Figure 8:** Dataset Insect Sound - Espèce *aedes female*

Enregistrements audios des sons émis par différentes espèces de mouches, des segments de 10 ms échantillonnés à 6000 Hz.

Comprend 10 espèces d'insectes. Nos expériences ne portent que sur une seule : l'*aedes female*. Dataset réduit à 5000 signaux en apprentissage et 5000 signaux en test.

Le début et la fin des signaux ressemble à du bruit blanc. Dans nos expérimentations, nous intéressons à une fenêtre centrale de chaque signal.

# Results

---

	DILATE	MSE	SOFT DTW
<b>ECG5000</b>			
MSE	$0.457 \pm 0.026$	$0.403 \pm 0.030$	$1.169 \pm 0.328$
DTW	$1.850 \pm 0.032$	$2.114 \pm 0.052$	$1.860 \pm 0.143$
TDI	$0.458 \pm 0.026$	$0.556 \pm 0.041$	$1.981 \pm 0.750$

**Table 1:** Tableau comparatif des résultats de prédiction par loss et par dataset

	DILATE	MSE	SOFT DTW
<b>Traffic</b>			
MSE	$0.567 \pm 0.030$	$0.363 \pm 0.004$	$1.085 \pm 0.132$
DTW	$1.831 \pm 0.005$	$1.836 \pm 0.010$	$1.836 \pm 0.080$
TDI	$0.408 \pm 0.023$	$0.402 \pm 0.003$	$2.415 \pm 0.467$

**Table 2:** Tableau comparatif des résultats de prédiction par loss et par dataset

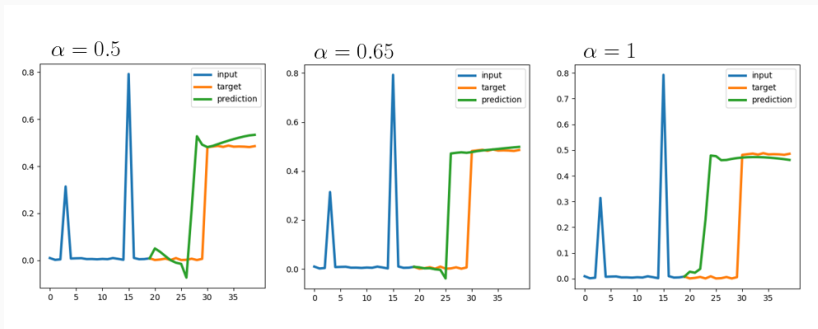
	DILATE	MSE	SOFT DTW
<b>Synth</b>			
MSE	$0.01689 \pm 0.001$	$0.0128 \pm 0.001$	$0.02117 \pm 0.00367$
DTW	$0.2083 \pm 0.0402$	$0.2633 \pm 0.0128$	$0.1517 \pm 0.0022$
TDI	$1.5688 \pm 0.2115$	$1.7504 \pm 0.1241$	$1.8313 \pm 0.0603$

**Table 3:** Tableau comparatif des résultats de prédiction par loss et par dataset

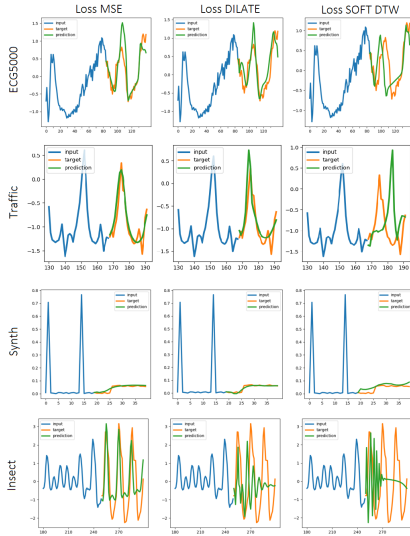


	DILATE	MSE	SOFT DTW
<b>Insect</b>			
MSE	$1.187 \pm 0.058$	$0.817 \pm 0.00049$	$1.185 \pm 0.077$
DTW	$3.361 \pm 0.052$	$3.568 \pm 0.083$	$3.213 \pm 0.0997$
TDI	$3.259 \pm 0.411$	$4.778 \pm 0.07$	$6.645 \pm 0.265$

**Table 4:** Tableau comparatif des résultats de prédiction par loss et par dataset



**Figure 9:** Influence du paramètre  $\alpha$  sur une série temporelle du dataset synthétique



# Conclusion

---



V. Le Guen and N. Thome.

**Shape and time distortion loss for training deep time series forecasting models.**

In *Advances in Neural Information Processing Systems*, pages 4191–4203. 2019.