Improved Learning via k-DTW: A Novel Dissimilarity Measure for Curves

Amer Krivošija, Alexander Munteanu,
André Nusser and Chris Schwiegelshohn

ICML 2025

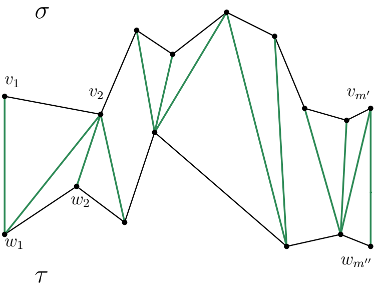
Motivation

Handwriting, panel data, time series, . . .
$\hat{=}$ polygonal curves in $\mathbb{R}^d$.

Comparing two curves requires a suitable dissimilarity measure.

Gold standards: the Fréchet and the Dynamic time warping distances.

Traversals



Given curves $\sigma = (v_1, \ldots, v_{m'})$ and $\tau = (w_1, \ldots, w_{m''})$ in $\mathbb{R}^d$.

A traversal T of σ and τ is a sequence of pairs of indices $T = (1, 1), \ldots, (m', m'')$, s.t.

$(i, j) \rightarrow (k, l) \in \{(i+1, j), (i, j+1), (i+1, j+1)\}$

k-DTW Definition

Let $\mathcal{T}$ be the set of all traversals T of σ and τ. Let the pair $(i, j) \in T$ attain the l-th largest distance $s_l^{(T)} = \|v_i - w_j\|$ in T s.t. $s_1^{(T)} \geq s_2^{(T)} \geq \ldots \geq s_{|T|}^{(T)}$. For any $l > |T|$ let $s_l^{(T)} = 0$.

Given $k \in \mathbb{N}$, the k-DTW distance of σ and τ is defined as

$$d_{k\text{-}DTW}(\sigma, \tau) = \min_{T \in \mathcal{T}} \sum_{l=1}^{k} s_l^{(T)}.$$

k-DTW generalizes both: discrete Fréchet and DTW distances.

Comparison to Standard Distances

Given curves $\sigma = (v_1, \ldots, v_{m'})$ and $\tau = (w_1, \ldots, w_{m''})$. Let $m = \max\{m', m''\}$.

Distance $\rightarrow$	Discrete Fréchet	Dynamic Time Warping	The new k-DTW
Triangle inequality	Yes	Factor m	Factor $k \ll m$
Robustness	Bad	Good	Good $k \gg 1$
Time to compute	$\tilde{\Theta}(m^2)$	$\tilde{\Theta}(m^2)$	$\tilde{\Omega}(m^2)$, ??, $O(m^4)$
$\ldots \varepsilon$-approximate	$\tilde{\Theta}(m^2)$	$\tilde{\Theta}(m^2)$	$\tilde{\Theta}(m^2)$

Think of $1 \ll k \ll m$, e.g., $k \in \Theta(\log m)$.

k-DTW Algorithms

Theorem: Exact algorithm

Given curves σ and τ, $|\sigma| = m'$ and $|\tau| = m''$, and a parameter k. We can compute $d_{\textit{k-DTW}}(\sigma, \tau)$ in time $O(m'm''z)$, where z is the number of distinct distances between vertices in σ and τ.

Problem:

z can be as large as $m'm'' \Rightarrow$ the running time is $O(m^4)$ in the worst case.

k-DTW Algorithms

Theorem: Exact algorithm

Given curves σ and τ, $|\sigma| = m'$ and $|\tau| = m''$, and a parameter k. We can compute $d_{k\text{-}DTW}(\sigma, \tau)$ in time $O(m'm''z)$, where z is the number of distinct distances between vertices in σ and τ.

Problem:
z can be as large as $m'm'' \Rightarrow$
the running time is $O(m^4)$ in the worst case.

Lemma: k-approximation

Given curves σ, τ, and a parameter k. Then $d_{dF}(\sigma, \tau)$ is a k-approximation for $d_{k\text{-}DTW}(\sigma, \tau)$. It holds that $d_{dF}(\sigma, \tau) \leq d_{k\text{-}DTW}(\sigma, \tau) \leq k \cdot d_{dF}(\sigma, \tau)$.

Theorem: $(1 + \varepsilon)$-approximation

Given curves σ and τ, $|\sigma| = m'$ and $|\tau| = m''$, and a parameter k. We can compute a $(1 + \varepsilon)$-approximation for k-DTW for any $0 < \varepsilon \leq 1$ in time

$$O\left(m'm''\frac{\log(k/\varepsilon)}{\varepsilon}\right) = \tilde{O}(m^2).$$

Improved Learning Theory via k-DTW

Problem: Learning the median of a distribution $\mathcal{D}$ over curves of complexity m supported in the Euclidean ball in $\mathbb{R}^d$.

$$\text{cost}(\mathcal{D}, \psi) := \int_\sigma d(\sigma, \psi)\mathbb{P}[\sigma]\ d\sigma$$

The excess risk of ψ_P on $P \sim \mathcal{D}^n$: $\mathcal{E} := \text{cost}(\mathcal{D}, \psi_P) - \min_\psi \text{cost}(\mathcal{D}, \psi)$ can be bounded by Rademacher or Gaussian complexities [Bartlett and Mendelson, 2002].

The Rademacher and Gaussian complexities for learning the median curve of complexity m are bounded above by

$$\tilde{O}\left(\sqrt{\frac{mk^2 \cdot \min\{d, k^2\}}{n}}\right)$$

under k-DTW and below by

$$\Omega\left(\sqrt{\frac{m^2}{n}}\right)$$

under DTW. This implies a separation when $k \in o(\sqrt{m/d})$ or $k \in o(m^{1/4})$.

Talk To Us

Please visit us during Poster Session 3 @ ICML 2025

Wednesday, July 16, 2025: 11:00a.m.-1:30p.m. PDT