基于保序最优传输的序列距离

Bing Su, Gang Hua

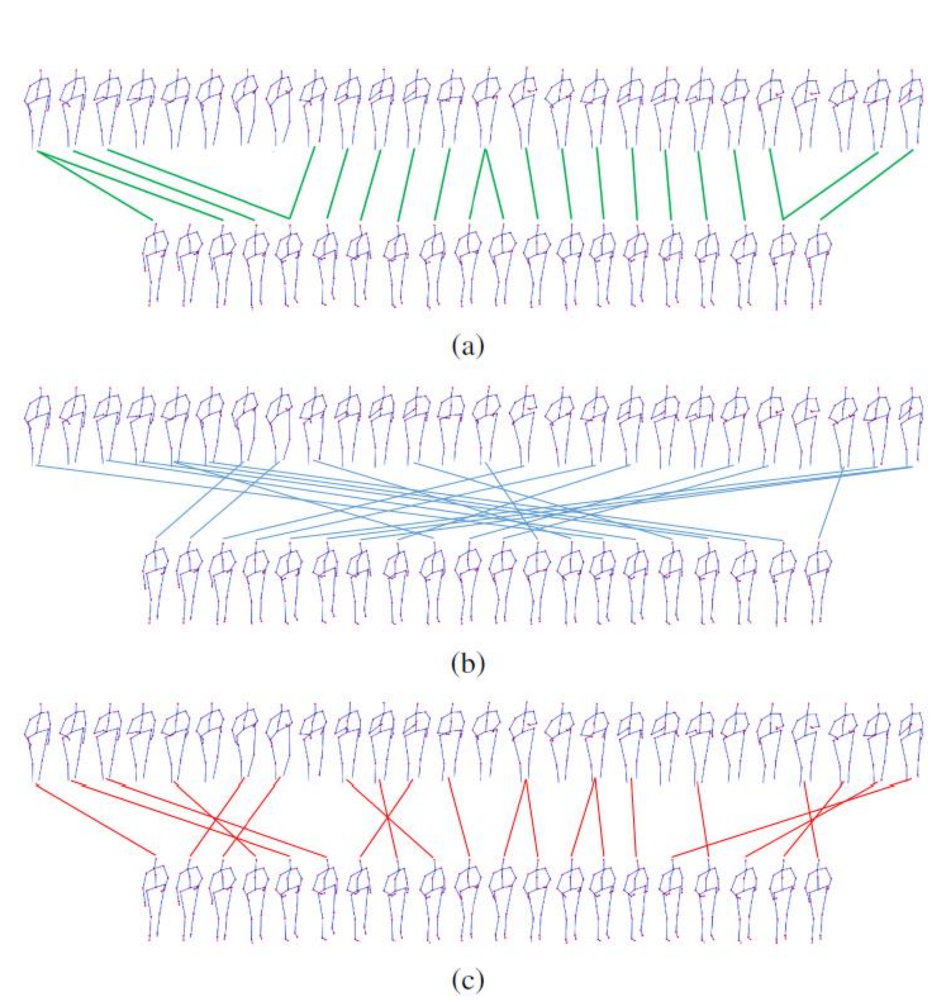
“Order-preserving Optimal Transport for Distances between Sequences”

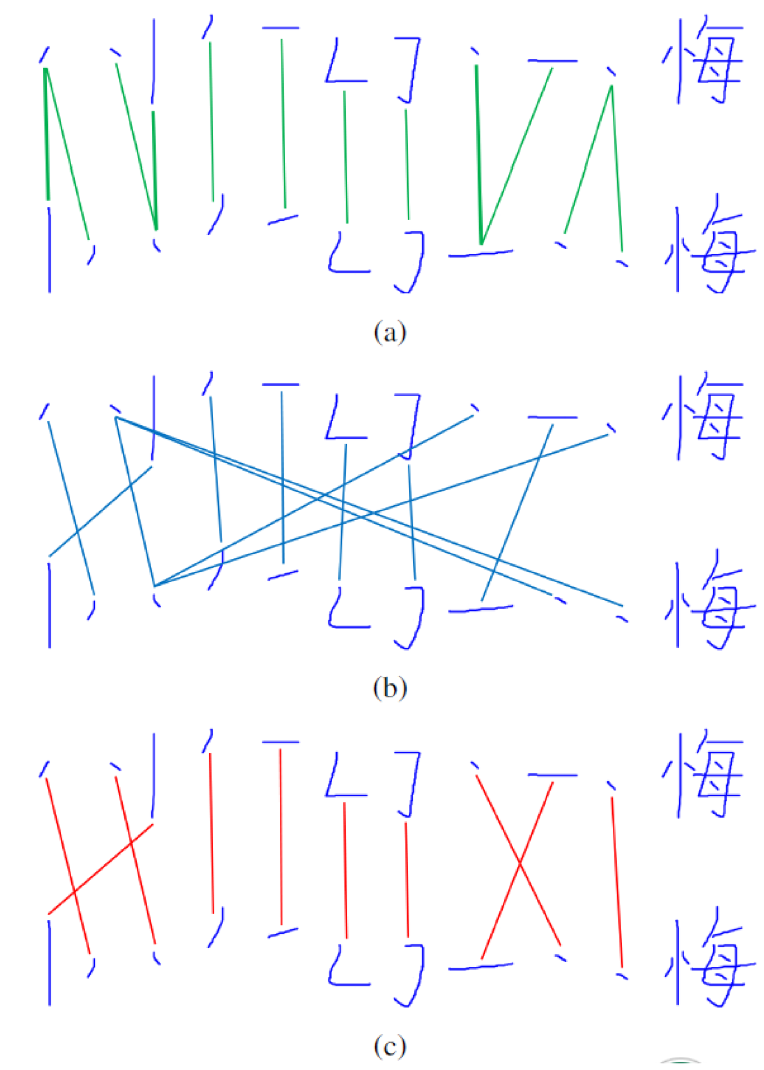
IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018.

苏冰，13661284169，subingats@gmail.com

• Motivation

- **Measuring the distance between sequences plays a fundamental role in sequence analysis problems**.
  - can be much more difficult than for vectors.
- **Various evolution speed & sampling rate**:
  - Different sequences may have different numbers of instances.
- **Evolutions are not uniform**:
  - Temporal alignments are necessary.
- **Instances in the same sequence are not independent**:
  - Instances are temporally related. Order-preserving is required.
- **Local disorders may exist**:
  - Strict order preservation may not be imposed to the alignment.
- **Arbitrary starting point for periodic patterns**





• Order-Preserving Wasserstein Distance

- **Cast alignment as the optimal transport problem**
  - Given two sequences $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N]$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_M]$, apply the optimal transport (OT) by viewing instances as independent samples: $d_O(\boldsymbol{X},\boldsymbol{Y}) = \min_{\boldsymbol{T}\in U(\boldsymbol{\alpha},\boldsymbol{\beta})} \langle \boldsymbol{T}, \boldsymbol{D}\rangle$

$$U(\boldsymbol{\alpha},\boldsymbol{\beta}) := \{\boldsymbol{T} \in \mathbb{R}_+^{N\times M} | \boldsymbol{T}\boldsymbol{1}_M = \boldsymbol{\alpha}, \boldsymbol{T}^T\boldsymbol{1}_N = \boldsymbol{\beta}\}$$

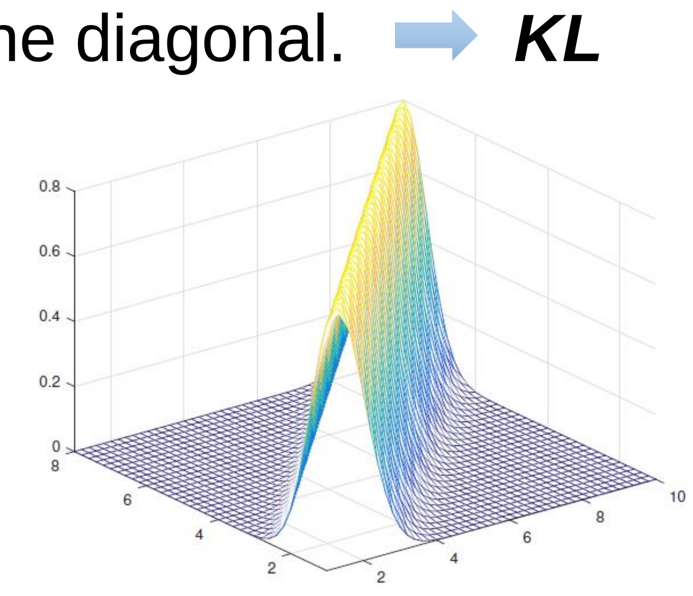
  - Problem: the ordering relationship is totally ignored.
- **Temporal order preserving regularizations**
  - Preserve the inherent temporal relationships of the instances
  - Desired: elements in one sequence are transported into those in the other sequence with similar relative temporal positions
  - The transport $\boldsymbol{T}$ should show local homogeneous structures: large values appear in the area near the diagonal, while the values in other areas are zero or very small. ➡ ***inverse difference moment (IDM)*** regularization

$$I(\boldsymbol{T}) = \sum_{i=1}^{N}\sum_{j=1}^{M} \frac{t_{ij}}{\left(\frac{i}{N} - \frac{j}{M}\right)^2 + 1}$$

  - A general ideal distribution of the values in $\boldsymbol{T}$ is that the peaks appear on the diagonal, and the values decrease gradually along the direction perpendicular to the diagonal. ➡ ***KL divergence*** with the Prior distribution

$$KL(\boldsymbol{T}||\boldsymbol{P}) = \sum_{i=1}^{N}\sum_{j=1}^{M} t_{ij} \log\frac{t_{ij}}{p_{ij}}$$

$$p_{ij} := \boldsymbol{P}(i,j) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\ell^2(i,j)}{2\sigma^2}}$$



- **Objective**
  - Adding the regularization terms to the OT objective

$$\min_{\boldsymbol{T}\in\mathbb{R}_+^{N\times M}} \langle \boldsymbol{T}, \boldsymbol{D}\rangle - \lambda_1 I(\boldsymbol{T}) + \lambda_2 KL(\boldsymbol{T}||\boldsymbol{P})$$
$$s.t.\ \boldsymbol{T}\boldsymbol{1}_M = \boldsymbol{\alpha}, \boldsymbol{T}^T\boldsymbol{1}_N = \boldsymbol{\beta}$$

- **Optimization**
  - By setting the Lagrangian function $\frac{\partial L(\boldsymbol{T},\boldsymbol{\mu},\boldsymbol{\nu})}{\partial t_{ij}}$ to 0, obtain $\boldsymbol{T}^{\lambda_1,\lambda_2} = \boldsymbol{e}^{diag(-\frac{1}{2}-\frac{\boldsymbol{\mu}}{\lambda_2})}\boldsymbol{K}\boldsymbol{e}^{diag(-\frac{1}{2}-\frac{\boldsymbol{\nu}}{\lambda_2})}$

$$s_{ij}^{\lambda_1} = \frac{\lambda_1}{(\frac{i}{N}-\frac{j}{M})^2+1} \qquad \boldsymbol{K} = [p_{ij} e^{\frac{1}{\lambda_2}(s_{ij}^{\lambda_1} - d_{ij})}]_{ij}$$

  - According to the Sinkhorn's theorem, the optimal transport T with this form is a rescaled version of K:
  - Efficiently solved by the Sinkhorn-Knopp iterative matrix scaling algorithm:

$$\boldsymbol{\kappa}_1 \leftarrow \boldsymbol{\alpha}./\boldsymbol{K}\boldsymbol{\kappa}_2 \qquad \boldsymbol{\kappa}_2 \leftarrow \boldsymbol{\beta}./(\boldsymbol{K})^T\boldsymbol{\kappa}_1$$

• Experimental Results

- **Comparison with other sequence distances**

Distance	DTW	IDTW	nDTW	Sinkhorn	OPW
Accuracy	71.06	73.63	70.70	66.67	74.36
MAP	50.77	58.05	56.55	51.43	59.10

(a) Results with the NM classifier

Distance	DTW	IDTW	nDTW	Sinkhorn	OPW
MAP	58.95	56.67	56.52	54.58	58.70
1-NN	81.32	82.78	79.85	78.02	84.25
3-NN	81.32	82.05	79.12	77.66	82.78
5-NN	80.95	79.12	76.92	74.73	80.22
15-NN	82.78	75.82	76.19	69.96	77.29

(b) Results with the NN classifier

Table 1. Results on the MSR Sports Action3D dataset. The best results among all distance measures are shown in red, and the second position results are shown in blue.

Method	DTW	IDTW	nDTW	Sinkhorn	OPW
Accuracy	40.63	33.75	38.75	37.50	38.75
MAP	33.25	37.47	31.37	32.14	33.35

(a) Results with the NM classifier

Distance	DTW	IDTW	nDTW	Sinkhorn	OPW
MAP	33.79	28.81	30.56	30.66	34.62
1-NN	58.75	50.00	55.63	54.37	58.13
3-NN	50.62	43.75	47.50	48.13	50.62
5-NN	49.38	50.00	52.50	50.62	53.75
15-NN	43.13	38.75	40.00	41.25	44.37

(b) Results with the NN classifier

Table 2. Results on the MSR Daily Activity3D dataset.

Method	DTW	IDTW	nDTW	Sinkhorn	OPW
Accuracy	45.01	46.10	41.31	34.97	57.75
	(2.21)	(3.92)	(3.40)	(1.84)	(2.62)
MAP	40.02	40.50	35.21	28.76	47.40
	(1.75)	(2.76)	(2.07)	(1.12)	(2.64)

(a) Results with the NM classifier

Distance	DTW	IDTW	nDTW	Sinkhorn	OPW
MAP	28.27	27.09	23.87	19.41	32.12
	(0.30)	(0.40)	(0.34)	(0.32)	(1.14)
1-NN	62.44	66.08	60.31	57.54	72.25
	(3.03)	(2.31)	(1.97)	(3.34)	(3.00)
3-NN	59.95	69.78	63.02	57.97	73.62
	(2.40)	(1.84)	(1.89)	(1.62)	(2.01)
5-NN	62.36	69.98	64.39	60.53	76.99
	(2.18)	(2.16)	(2.31)	(2.04)	(2.01)
15-NN	63.77	72.19	64.24	61.29	78.38
	(1.76)	(4.33)	(2.77)	(2.49)	(2.49)

(b) Results with the NN classifier

Table 3. Results on similar online Chinese character recognition. Standard deviations are shown in brackets.

Method	DTW	IDTW	nDTW	Sinkhorn	OPW
Accuracy	82.55	80.18	76.27	77.00	87.27
MAP	73.23	76.29	67.75	65.11	82.23

(a) Results with the NM classifier

Distance	DTW	IDTW	nDTW	Sinkhorn	OPW
MAP	56.58	56.03	48.47	43.27	62.71
1-NN	96.36	96.73	95.05	87.95	96.68
3-NN	96.91	96.82	95.73	89.05	97.45
5-NN	97.23	96.73	96.09	89.23	97.14
15-NN	97.36	96.50	95.91	90.73	97.41

(b) Results with the NN classifier

Table 4. Results on the SAD dataset.