

A study of sequence matching method considering data transition

Masashi Kamura^{*1} Hiroyuki Kasai^{*1,*2}

^{*1} Department of Computer Science and Communications Engineering,
Graduate School of Fundamental Science and Engineering, Waseda University

^{*2} Department of Communications and Computer Engineering,
School of Fundamental Science and Engineering, Waseda University

This paper addresses a problem of measuring a distance between sequences of which order has some meanings. One representative method is Order-Preserving Wasserstein distance (OPW). OPW has a disadvantage that it does not fully take into account the similarity and neighbor relationship of the elements. To this end, building upon the framework of OPW, we propose a method that considers the neighbor relationship between elements by using the transitions of elements. From numerical evaluation experiments, we show that the proposed method improves classification accuracies across some real-world datasets.

1. Introduction

Sequence data or time series data refers to data that changes over time and has meanings in its order [Narimatsu 15, Narimatsu 17, Narimatsu 20]. The similarity between two sequence data is calculated based on the distance between these two sequences. *Sequence matching* is used to calculate such distances, which is the correspondence between the elements with these two sequences. The correspondences and distances obtained by sequence matching can be used in a wide range of fields. When performing sequence matching, the following four points should be considered: (i) similarity for each elements, (ii) order of the elements, (iii) relative temporal position, and (iv) neighbor relationship of the elements.

Dynamic Time Warping (DTW) [Sakoe 78] is widely used for sequence matching, and it has been extensively studied to enhance its capabilities and flexibilities [Jeong 11, Keogh 01, Cuturi 18, Zhao 16]. However, DTW has a problem that does not handle the order difference. Recently, OPW [Su 17] has been proposed as a sequence matching method using a transport plan calculated by the optimal transport (OT) problem [Villani 08, Peyré 19]. Optimal transport regards each element as a sample in a probability distribution, and calculates the optimal transport plan between them. It has been widely used in machine learning because it can express Wasserstein distance [Fukunaga 20, Huang 21, Kasai 20, Fukunaga 22]. In addition to the transport plan considered in OT, OPW imposes the constraint that the elements that are temporally close should be mutually matched, but it does not take into account the item (iv) mentioned above. Therefore, in this paper, we propose a sequence matching method that extends OPW to take into account the neighbor relationship of elements. In the numerical evaluation experiments, we conduct classification experiments on real-world

datasets using the nearest neighbor classifier, and show that the proposed matching method can yield better classification performances than the existing methods do. Hereinafter, we present vectors as \mathbf{a}, \mathbf{b} , and matrices as \mathbf{A}, \mathbf{B} . The Frobenius inner product for matrices \mathbf{A} and \mathbf{B} which have the same size is represented as $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} \mathbf{A}_{i,j} \mathbf{B}_{i,j}$.

Also, sequence data is represented as a matrix with N elements, where each element is a vector of dimension d , i.e., $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$.

2. Proposed Method

As mentioned in **Section 1**, OPW does not fully consider neighbor relationship of elements because it performs matching that strongly considers the relative temporal position. Therefore, it is necessary to achieve matching between elements that are distant but similar in their relative time-series locations. To this end, this paper focuses on the prior distribution \mathbf{P} in OPW, and proposes a similarity distribution based on the transition of the elements. To tackle the problem above, we consider to group sequence and perform sequence matching by OPW between grouped subsequences. This is achieved based on the relationship with neighbor elements.

More concretely, as a first step, the relationship of each element to its neighbors, denoted as $d(\mathbf{x}_k, \mathbf{x}_{k-1})$, is represented as follows:

$$d(\mathbf{x}_k, \mathbf{x}_{k-1}) = \|\mathbf{x}_k - \mathbf{x}_{k-1}\|_2.$$

Note that, for the first element \mathbf{x}_1 , we use its ℓ_2 norm as $\|\mathbf{x}_1\|_2$. Next, in order to consider the degree of variation in the similarity between the previous and next elements as a whole sequence, the relative element transitions are represented as follows:

$$f(\mathbf{X})_i = \frac{\sum_{k=1}^i d(\mathbf{x}_k, \mathbf{x}_{k-1})}{\sum_k d(\mathbf{x}_k, \mathbf{x}_{k-1})} (i = 1, 2, \dots, N). \quad (1)$$

Contact: Masashi Kamura, Department of Computer Science and Communications Engineering, Graduate School of Fundamental Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan, kamura610@ruri.waseda.jp

From equation (1), we define the similarity distribution \mathbf{P} which considers the changes in transition of the two sequences as follows:

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

where $\ell(i, j)$ is calculated in the following two formulas:

$$(I) \ell(i, j) = \begin{cases} 1, & (f(\mathbf{X})_i = f(\mathbf{Y})_j) \\ \frac{\max\{f(\mathbf{X})_i, f(\mathbf{Y})_j\}}{\min\{f(\mathbf{X})_i, f(\mathbf{Y})_j\}}, & (otherwise) \end{cases} \quad (2)$$

$$(II) \ell(i, j) = \frac{|f(\mathbf{X})_i - f(\mathbf{Y})_j|}{\sqrt{1/N^2 + 1/M^2}}. \quad (3)$$

Accordingly, the optimal matching matrix \mathbf{T}^* is given by following the formulation of OPW:

$$\mathbf{T}^* = \arg \min_{\mathbf{T} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \langle \mathbf{T}, \mathbf{D} \rangle - \lambda_1 I(\mathbf{T}) + \lambda_2 KL(\mathbf{T} \| \mathbf{P}),$$

where \mathbf{T} is matching matrix, \mathbf{U} is transportation polytope, \mathbf{a}, \mathbf{b} is a sample in a probability distribution for each sequence, $I(\mathbf{T})$ is the inverse difference moment, and $KL(\cdot \| \cdot)$ is the Kullback-Leibler divergence. Finally, the distance between two sequences is defined as:

$$\text{dist}(\mathbf{X}, \mathbf{Y}) = \langle \mathbf{T}^*, \mathbf{D} \rangle.$$

In this paper, we designate the proposed method using (I) in (2) as Transition OPW-I (TOPW-I). We also denote the one using (II) in (3) as TOPW-II.

3. Evaluations

This section presents visualizations of the effect of the proposed distance-based sequence matching. We also conducts classification experiments using the nearest neighbor (1-NN) classifier. The accuracy is used as the evaluation metric. For the former experiments, **a generated synthetic sequence data is used, and we show the matching matrix between them.** The datasets used in the classification experiments are Spoken Arabic Digit (SAD) [Asuncion 07]^{*1}, BME, CBF, and Fish from UCR Time Series Classification Archive [Dau 19]^{*2}. The details of the datasets are summarized in Table 1.

Table 1: Detail of datasets

dataset	# class	train	test	dim	length
SAD	10	6600	2200	13	indefinite
BME	3	30	150	1	128
CBF	3	30	900	1	128
Fish	7	175	175	1	463

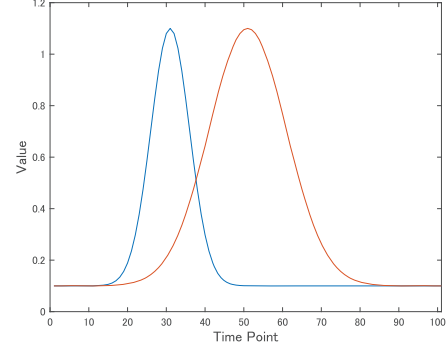


Figure 1: data

3.1 Visualizations of Matching Effect

We verify whether the proposed matching method can achieve matching between elements whose temporal positions are close but similar. For this purpose, we verified this by visualizing the matching matrix \mathbf{T}^* when sequence matching is performed using 1-dimensional sequence data. **The sequence data used for this evaluation is shown in Figure 1.** The blue line is a sample sequence, and the orange one is **a sequence that differs from the blue one** in terms of the timing to reach the peak and the fluctuation. Both sequence data have similar characteristics in that the values rise from a flat state, reach a peak, and then fall down to a flat state. Therefore, we expect that the flat parts or the rising parts of the two sequence data are matched each other. The sequence matching between these sequence data by each method is shown in Figure 2. The closer to yellow it is, the larger the higher the matching probability is.

3.2 Classification Evaluations

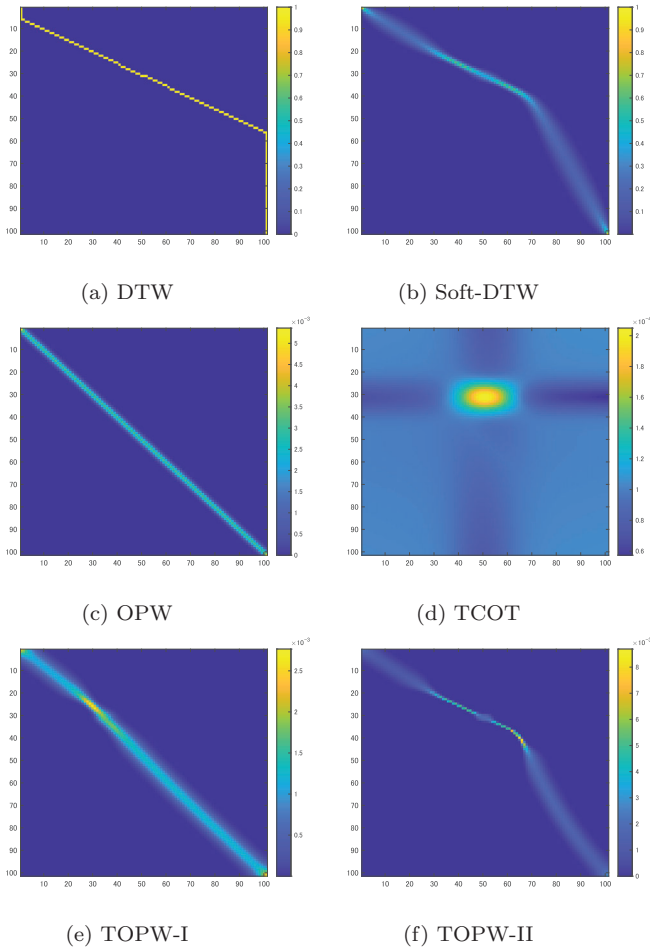
The classification results using 1-NN classifier based on the distance between two sequences are shown in Table 2. The best results are indicated in **bold**. The comparison methods are DTW, Soft-DTW [Cuturi 18] which is the variation of DTW, OPW which is the basis of our proposed method, and TCOT [Su 19] which is similar with OPW. We set $(\lambda_1, \lambda_2, \sigma) = (50, 0.1, 1)$ for OPW and the proposed TOPW. We also set $\gamma = 0.1$ for Soft-DTW, and set $\lambda = 50$ for TCOT. From Table 2, our proposed TOPWs show better performances over several datasets.

Table 2: Classification Accuracy by 1-NN.

	SAD	BME	CBF	Fish
DTW	0.9636	0.8933	0.9967	0.8229
Soft-DTW	0.9623	0.9733	0.9767	0.1257
OPW	0.9636	0.8600	0.9889	0.7829
TCOT	0.8505	0.6667	0.7022	0.7429
TOPW-I	0.9659	0.9467	0.9967	0.8571
TOPW-II	0.9627	0.9933	0.9722	0.7886

*1 <https://archive.ics.uci.edu/ml/datasets/Spoken+Arabic+Digit>

*2 https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

Figure 2: Visualization of \mathbf{T}^* in each method.

4. Conclusion

We have proposed a method to calculate the transition of how much it varied in the whole sequence, and used it to generate a similar distribution. Through the visualization evaluation, the synthetic dataset confirmed that the proposed method captured the fluctuation and performed appropriate matching. The numerical experiments by sequence classification show that the method outperforms existing methods across several datasets.

References

- [Asuncion 07] Asuncion, A. and Newman, D.: UCI machine learning repository (2007)
- [Cuturi 18] Cuturi, M. and Blondel, M.: Soft-DTW: a Differentiable Loss Function for Time-Series, *arXiv preprint arXiv:1703.01541* (2018)
- [Dau 19] Dau, H. A., Bagnall, A., Kamgar, K., Yeh, C.-C. M., Zhu, Y., Gharghabi, S., Ratanamahatana, C. A., and Keogh, E.: The UCR Time Series Archive, *arXiv preprint arXiv:1810.07758* (2019)
- [Fukunaga 20] Fukunaga, T. and Kasai, H.: Wasserstein k -means with sparse simplex projection, in *ICPR* (2020)
- [Fukunaga 22] Fukunaga, T. and Kasai, H.: Block-coordinate Frank–Wolfe algorithm and convergence analysis for semi-relaxed optimal transport problem, in *ICASSP* (2022)
- [Huang 21] Huang, J., Fang, Z., and Kasai, H.: LCS graph kernel based on Wasserstein distance in longest common subsequence metric space, *Digital Signal Processing*, Vol. 189, p. 108281 (2021)
- [Jeong 11] Jeong, Y.-S., Jeong, M. K., and Omitaomu, O. A.: Weighted dynamic time warping for time series classification, *Pattern recognition*, Vol. 44, No. 9, pp. 2231–2240 (2011)
- [Kasai 20] Kasai, H.: Multi-view Wasserstein discriminant analysis with entropic regularized Wasserstein distance, in *ICASSP* (2020)
- [Keogh 01] Keogh, E. J. and Pazzani, M. J.: Derivative dynamic time warping, in *SDM*, pp. 1–11 (2001)
- [Narimatsu 15] Narimatsu, H. and Kasai, H.: Duration and Interval Hidden Markov Model for Sequential Data Analysis, *IJCNN* (2015)
- [Narimatsu 17] Narimatsu, H. and Kasai, H.: State duration and interval modeling in hidden semi-Markov model for sequential data analysis, *Annals of Mathematics and Artificial Intelligence*, Vol. 81, No. 3-4, pp. 377–403 (2017)
- [Narimatsu 20] Narimatsu, H. and Kasai, H.: Overlapped State Hidden Semi-Markov Model for Grouped Multiple Sequences, in *ICASSP* (2020)
- [Peyré 19] Peyré, G. and Cuturi, M.: Computational optimal transport, *Foundations and Trends in Machine Learning*, Vol. 11, No. 5-6, pp. 355–607 (2019)
- [Sakoe 78] Sakoe, H. and Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition, *IEEE Trans. Signal Process*, Vol. 26, No. 1, pp. 43–49 (1978)
- [Su 17] Su, B. and Hua, G.: Order-preserving Wasserstein distance for sequence matching, in *CVPR* (2017)
- [Su 19] Su, B. and Hua, G.: Order-preserving optimal transport for distances between sequences, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 41, No. 12, pp. 2961–2974 (2019)
- [Villani 08] Villani, C.: *Optimal transport: old and new*, Springer Science & Business Media (2008)
- [Zhao 16] Zhao, J. and Itti, L.: shapeDTW: shape Dynamic Time Warping, *arXiv preprint arXiv:1606.01601* (2016)