

データの推移を考慮したシーケンスマッチング手法の検討

嘉村 雅志^{1,a)} 笠井 裕之^{1,2,b)}

概要: 本稿では、順序に意味を持つシーケンス間の距離を測定する問題について着目する。シーケンス間の距離を測定する上で、シーケンスの各要素の類似度、要素の順序、相対的な時系列の位置関係、隣接関係を考慮することが重要である。既存手法である Order-Preserving Wasserstein distance (OPW) では、隣接関係を考慮しないという問題点が存在する。そこで本稿では、要素間の隣接関係を考慮するために、OPW によるマッチングに、要素の推移を用いたマッチングを付加する手法を提案する。マッチングの様子の視覚的な評価から、提案手法が既存手法よりもデータの隣接関係を考慮したマッチングを行うことを確認する。また、数値評価実験から、いくつかのデータセットに対して、提案手法によりクラス分類精度が向上することを示す。

A study of sequence matching method considering data transition

1. はじめに

シーケンスデータ（時系列データ）とは、気温や株価の変動、映像、音声といった、時間とともに変化し、その順序に意味のあるデータのことを指す。2つのシーケンス間の類似度は、対象となる2つのシーケンス間の距離を基に計算される。この距離を計算する際に、シーケンスマッチングが用いられる。シーケンスマッチングとは、シーケンスを構成する各要素間の対応関係のことである。シーケンスマッチングにより求められる対応関係や距離は、時系列データの分類や音声認識、映像認識、映像予測といった幅広い分野に利用される。

シーケンスマッチングを行う際に考慮すべき事項は以下の4点である。

- (1) 各要素ごとの類似度
- (2) 要素の順序
- (3) 相対的な時系列の位置関係
- (4) 要素の隣接関係

各要素ごとの類似度は、シーケンスを構成する各要素がベクトルデータであるという仮定の下、ユークリッド距離や l_p ノルムといった、ベクトルデータ間の距離を用いて計算される。シーケンスマッチングにおいて、各要素ごとの

類似度を考慮することは自然である。しかし、順序に意味のあるシーケンスにおいて、要素が同じでその順序が異なるシーケンス、あるいは、展開の速さが異なるシーケンスのマッチングに対して、要素の類似度を考慮するだけでは不十分である。そこで、時系列において相対的に近い要素同士をマッチングさせる制約を課すことで、要素が同じでも順序の異なるシーケンスおよび展開の速さが異なるシーケンス間のマッチングに対して、順序の違いを反映させることができるようになる。さらに、隣接関係も考慮する必要がある。ある2つのシーケンスにおいて、類似した要素で時系列の相対的な位置が近くても、その前後の要素が異なるものであるとき、類似した要素はマッチングするべきではない。例えば、「座って立つ」という動作は、「座る」「立つ」という動作で構成されているが、各動作の長さはシーケンスにより異なる。このとき、「座る直前」を表す要素と「立つ直前」を表す要素同士がマッチングする可能性があるが、この2つの要素が持つ意味は異なるため、不適切なマッチングである。以上を考慮して適切にマッチングさせることが必要である。

近年、シーケンスの各要素を確率分布におけるサンプルとみなし、最適輸送問題 (Optimal Transport, OT) [1], [2] により算出される輸送計画を用いたシーケンスマッチング手法として、Order-Preserving Wasserstein distance (OPW) [3], [4] が提案されている。OPW では、OT による輸送計画に加えて、時系列の相対的に近い要素同士が強くマッ

¹ 早稲田大学大学院基幹理工学研究科情報理工・通信専攻

² 早稲田大学基幹理工学部情報通信学科

^{a)} kamura610@ruri.waseda.jp

^{b)} hiroyuki.kasai@waseda.jp

ングする制約を課しているが、上述の (4) を考慮していない。そこで、本稿では、OPW を拡張し、要素の隣接関係を考慮したシーケンスマッチング手法を提案する。

数値評価実験では、最近傍法による実世界データセットの分類実験を行い、提案手法によるマッチングが、既存手法と比較して優れた分類を行うことができることを示す。

2. 関連研究

表記法 本稿では、ベクトルは \mathbf{a}, \mathbf{b} のように表され、行列は \mathbf{A}, \mathbf{B} のように表される。同じサイズの行列 \mathbf{A}, \mathbf{B} に対するフロベニウス内積は、 $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} \mathbf{A}_{i,j} \mathbf{B}_{i,j}$ と表される。また、シーケンスデータは、各要素が d 次元のベクトルで要素数 N の行列、つまり、 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$ で表される。

2.1 動的時間伸縮法 (Dynamic Time Warping)

シーケンスマッチングにおける代表的な手法として、動的時間伸縮法 (Dynamic Time Warping, DTW) [5] が提案されており、関連研究も数多く存在する [6], [7], [8]。DTW では、各要素間のユークリッド距離を基に、以下の制約の下でマッチング後の距離が最小となるマッチングを算出する。

- (1) 境界条件
- (2) 単調性
- (3) 連続性

DTW では、上記の制約から、順序の異なる要素を持つシーケンス間では誤ったマッチングを行ってしまうという問題点が存在する。

2.2 最適輸送 (Optimal Transport) 問題と Order-Preserving Wasserstein distance (OPW)

Order-Preserving Wasserstein distance (OPW) は、OT により算出される輸送計画を用いたシーケンスマッチング手法である。本稿で提案する手法は OPW を拡張するため、OPW の概要を説明する。

2つのシーケンス $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_M]$ 間の OPW による距離は、以下で定義される。

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

ここで、 $\mathbf{D} = [d_{ij}] \in \mathbb{R}^{N \times M}$ は各要素間の距離を表す距離行列、 $\mathbf{T}^* = [t_{ij}] \in \mathbb{R}^{N \times M}$ は最適な輸送行列であり、それぞれ以下のように表される。

$$\mathbf{D}_{i,j} = d(\mathbf{x}_i, \mathbf{y}_j)^p,$$

$$\mathbf{T}^* = \arg \min_{\mathbf{T} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \langle \mathbf{T}, \mathbf{D} \rangle - \lambda_1 I(\mathbf{T}) + \lambda_2 KL(\mathbf{T} \parallel \mathbf{P}). \quad (1)$$

ただし、 \mathbf{T} は輸送行列、 $\lambda_1, \lambda_2 > 0$ は正則化パラメータである。また、 $I(\mathbf{T})$ は \mathbf{T} の逆差分モーメント、 $KL(\cdot \parallel \cdot)$ は

カルバック・ライブラー情報量を表している。逆差分モーメントは以下のように定義される。

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

さらに、 \mathbf{P} は事前分布であり、以下のように定義される。

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

$$\ell(i,j) = \frac{|i/N - j/M|}{\sqrt{1/N^2 + 1/M^2}}.$$

OT において、 t_{ij} が輸送量、 d_{ij} が輸送コストを表し、 d_{ij} が小さいほど t_{ij} の値が大きくなる。シーケンスマッチングにおいては、輸送量を要素間のマッチング確率、輸送コストを各要素間の類似度とみなす。つまり、2つのシーケンス間の OT を求めることは、類似した要素間のマッチング確率が高くなるようなマッチングを求めることと同義である。そこで、本稿では、輸送行列 \mathbf{T} をマッチング行列と呼ぶこととする。また、式 (1) の第 2 項により、マッチング行列 \mathbf{T} の対角成分の値を大きくし、第 3 項により、マッチング行列 \mathbf{T} を事前分布 \mathbf{P} に近づける。この 2つの正則化により、時系列の相対的に近い要素同士のマッチングを促す。

3. 提案手法

3.1 OPW の問題点と提案手法のアイデア

OPW では、時系列の相対的な位置関係を強く考慮したマッチングを行うため、隣接関係を考慮していないという問題点がある。そこで本稿では、OPW における事前分布 \mathbf{P} に着目し、隣接関係を考慮した事前分布にすることを考える。具体的には、シーケンスごとの要素の推移を計算し、それを基に類似分布を計算する。

3.2 提案手法のアプローチ

最初に、各シーケンスごとに、隣接関係の類似度を以下で計算する。

$$d_t(\mathbf{x}_k, \mathbf{x}_{k-1}) = \|\mathbf{x}_k - \mathbf{x}_{k-1}\|_2. \quad (2)$$

ただし、最初の要素については、最初の要素の ℓ_2 ノルムとする。

次に、相対的な要素の推移を以下のように表す。

$$f(\mathbf{X})_i = \frac{\sum_{k=1}^i d_t(\mathbf{x}_k, \mathbf{x}_{k-1})}{\sum_k d_t(\mathbf{x}_k, \mathbf{x}_{k-1})} (i = 1, 2, \dots, N-1). \quad (3)$$

式 (3) を用いて、2つのシーケンスの推移の変化を考慮した類似分布 \mathbf{P} を以下のように定義する。

$$\mathbf{P}_{i,j} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\ell^2(i,j)}{2\sigma^2}}. \quad (4)$$

ここで、 ℓ について以下の2つを採用する。

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

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

そして、OPW の定式化に則って、マッチング行列を以下のように定義する。

$$\mathbf{T}^* = \arg \min_{\mathbf{T} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \langle \mathbf{T}, \mathbf{D} \rangle - \lambda_1 I(\mathbf{T}) + \lambda_2 KL(\mathbf{T} \| \mathbf{P}). \quad (5)$$

最終的に、2つのシーケンス間の距離は以下のように定義される。

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

アルゴリズムを以下に示す。

Algorithm 1 提案手法

Require: two sequences $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_M]$, parameter $\lambda_1, \lambda_2, \sigma$

Ensure: $dist(\mathbf{X}, \mathbf{Y})$, \mathbf{T}^*

- 1: Calculate the neighbor similarity by Eq.(2)
- 2: Calculate relative data transition by Eq.(3)
- 3: Calculate similarity distribution by Eq.(4)
- 4: **while** iteration has not reached to the maximum number **do**
- 5: Calculate matching matrix \mathbf{T}^* by Eq.(5)
- 6: **end while**
- 7: Calculate $dist(\mathbf{X}, \mathbf{Y}) = \langle \mathbf{T}^*, \mathbf{D} \rangle$

4. 実験

本章では、シーケンスデータとシーケンスマッチングの様子を可視化することによる視覚的評価と、提案手法により計算される距離を用いた最近傍法 (1-Nearest Neighbor, 1-NN), mean Average Precision (mAP) による分類実験の精度、処理時間による評価を行う。実験に用いたデータセットは、Spoke Arabic Digit (SAD) [9] *1, UCR Time Series Classification Archive [10] *2 より FaceAll, MedicalImages である。データセットの詳細を以下に示す。

表 1 データセットの詳細

データセット	クラス数	学習データ数	テストデータ数	次元	長さ
SAD	10	6600	2200	13	不定
FaceAll	14	560	1690	1	131
MedicalImages	10	381	760	1	99

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

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

4.1 視覚的評価

本節では、提案手法と、提案手法の基礎となる、既存手法の OPW を用いて、シーケンスデータ間のマッチングの様子を可視化することで視覚的評価を行う。本実験を行う上で用いたシーケンスデータを図 1 に示す。2つのシーケンスデータは同じクラスに属しており、ピークの数等は等しいが、ピークの起こるタイミングが異なるデータである。この2つのシーケンスデータ間の OPW によるシーケンスマッチングの様子を図 2 に示す。また、提案手法 (i) (ii) によるシーケンスマッチングの様子をそれぞれ図 3, 図 4 に示す。ここで、図 2, 図 3, 図 4 において、正方形の図がマッチング行列を表しており、黄色に近づくほど値が大きい、つまり、マッチング確率が高いことを示している。図 2 では、データの推移に関係なくマッチング行列の対角成分の値が大きくなっている。また、提案手法 (i) では、値の分布の幅が広がっているものの、依然対角成分の値が大きい結果となった。一方、提案手法 (ii) については、最初のピーク同士の部分でマッチング行列の値が大きくなるような結果となった。

4.2 分類実験

シーケンス間の距離を用いた最近傍法による分類実験の結果を表 2, 表 3 に示す。比較手法はシーケンスマッチングで広く用いられている DTW, 提案手法の基礎となる OPW である。OPW と提案手法におけるパラメータは $(\lambda_1, \lambda_2) = (50, 0.1), (5, 0.1)$ を用いる。また、 $\sigma = 1$ で固定した。最も良い結果を太字で、提案手法において OPW よりも良い結果を下線で示している。表 2 より、SAD では提案手法 (i) で 1-NN, mAP の値が最も大きい結果となった。また、SAD の提案手法 (ii), FaceAll の提案手法 (i) (ii), MedicalImages の提案手法 (i) で 1-NN の値が、全データセットの提案手法 (ii) で mAP の値が OPW よりも優れた性能を示した。表 3 より、MedicalImages では提

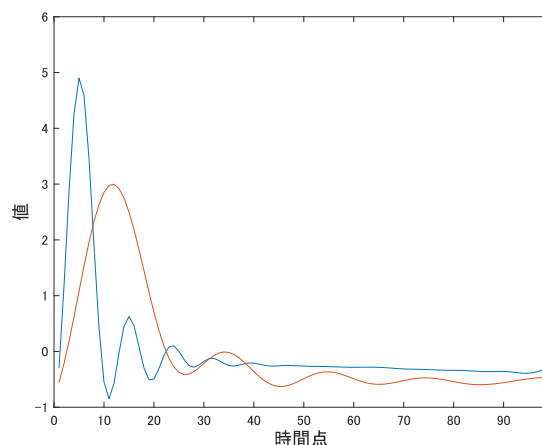


図 1 シーケンスデータ

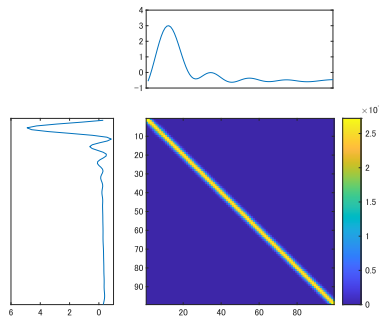


図 2 既存手法のマッチング

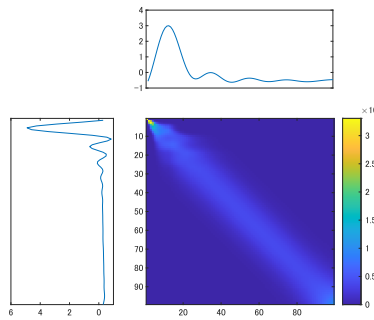


図 3 提案手法 (i) のマッチング

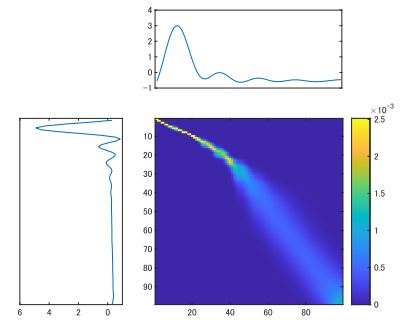


図 4 提案手法 (ii) のマッチング

表 2 最近傍法による分類精度 ($\lambda_1 = 50, \lambda_2 = 0.1$)

		DTW	OPW	提案手法 (1)	提案手法 (2)
SAD	1-NN	0.9636	0.9636	0.9659	0.9627
	mAP	0.5658	0.5977	0.6054	0.6006
	処理時間 [sec]	1.7856	0.9820	1.5540	1.4834
FaceAll	1-NN	0.8077	0.7130	0.7686	0.7639
	mAP	0.5483	0.3138	0.3008	0.4929
	処理時間 [sec]	1.5207	0.9549	1.2340	1.0804
MedicalImages	1-NN	0.7368	0.7039	0.7039	0.7342
	mAP	0.4844	0.4419	0.3826	0.4730
	処理時間 [sec]	0.5619	0.2637	0.4324	0.4158

表 3 最近傍法による分類精度 ($\lambda_1 = 5, \lambda_2 = 0.1$)

		DTW	OPW	提案手法 (1)	提案手法 (2)
SAD	1-NN	0.9636	0.9627	0.9614	0.9550
	mAP	0.5658	0.6071	0.5875	0.5995
	処理時間 [sec]	1.1378	0.8471	1.4668	1.4835
FaceAll	1-NN	0.8077	0.7118	0.7284	0.7503
	mAP	0.5483	0.3160	0.2526	0.4807
	処理時間 [sec]	1.5207	0.8654	1.3886	1.1010
MedicalImages	1-NN	0.7368	0.6961	0.6684	0.7395
	mAP	0.4844	0.4417	0.3688	0.4723
	処理時間 [sec]	0.6020	0.3042	0.4232	0.4345

案手法 (ii) で 1-NN の値が最も大きい結果となった。また、FaceAll の提案手法 (i) (ii) で 1-NN の値が、FaceAll, MedicalImages の提案手法 (ii) で mAP の値が OPW よりも優れた性能を示した。処理時間については、いずれも OPW が最も優れていた。これは、提案手法では、分布を各シーケンスごとに計算するのに対して、OPW では事前分布がシーケンスに依存しないことが原因であると考えられる。

5. まとめ

本稿では、最適輸送問題を用いたシーケンスマッチングにおける問題点を指摘し、それを改善する手法を提案した。具体的には、最適輸送問題を用いたシーケンスマッチング手法である OPW において、隣接関係を考慮していないという問題点があり、それに対して、事前分布を要素の推移を考慮したものにすることで、隣接関係を考慮することができるようにした。視覚的な評価実験から、提案手法において、ピークの起こる位置の異なるシーケンスデータに対

して適切なマッチングを実現した。また、数値実験から、一部のデータセットに対して既存手法よりも優れた性能を示した。今後は、現在類似分布の形で隣接関係を考慮していたものを、明示的に表すことを検討していく。

参考文献

- [1] Villani, C.: *Optimal transport: old and new*, Vol. 338, Springer Science & Business Media (2008).
- [2] Peyré, G. and Cuturi, M.: *Computational Optimal Transport, Foundations and Trends in Machine Learning*, Vol. 11, No. 5-6, pp. 355–607 (2019).
- [3] Su, B. and Hua, G.: Order-Preserving Wasserstein Distance for Sequence Matching, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
- [4] Su, B. and Hua, G.: Order-Preserving Optimal Transport for Distances between Sequences, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 12, pp. 2961–2974 (2019).
- [5] Sakoe, H. and Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition, *IEEE transactions on acoustics, speech, and signal processing*, Vol. 26, No. 1, pp. 43–49 (1978).
- [6] Jeong, Y.-S., Jeong, M. K. and Omiaomu, O. A.: Weighted dynamic time warping for time series classification, *Pattern recognition*, Vol. 44, No. 9, pp. 2231–2240 (2011).
- [7] Keogh, E. J. and Pazzani, M. J.: Derivative dynamic time warping, *Proceedings of the 2001 SIAM international conference on data mining*, SIAM, pp. 1–11 (2001).
- [8] Zhao, J. and Itti, L.: shapedtw: Shape dynamic time warping, *Pattern Recognition*, Vol. 74, pp. 171–184 (2018).
- [9] Asuncion, A. and Newman, D.: UCI machine learning repository (2007).
- [10] 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, *IEEE/CAA Journal of Automatica Sinica*, Vol. 6, No. 6, pp. 1293–1305 (2019).