

隣接要素の関連と時間分布の違いを考慮した グループ化シーケンスマッチング手法の検討

堀江 光彦^{1,a)} 笠井 裕之^{1,2,b)}

概要: 本稿では、シーケンスマッチングを用いたシーケンスデータ間の距離計算手法の提案を行う。映像や音声データもシーケンスデータに含まれるため、その距離計算は動作検知や会話認識、およびその学習などに利用することができる。本稿で提案する手法では、最適輸送問題を用いたマッチング手法を拡張し、マッチング情報を基に類似した隣接要素を判別しグループ化することを目標とする。類似した要素をグループ化し、グループ単位でマッチングし直すことで、シーケンスの時間分布の違いや隣接要素の関連を考慮したマッチングを実現する。さらに、提案手法について、既存手法と比較する数値実験を行い、提案手法の有効性を示す。

1. はじめに

本稿ではシーケンスデータ間の新たな距離計算手法を提案する。映像や音声といったデータはシーケンスデータに分類されるため、適切な距離計算を行うことは、動作検知や会話認識など、マルチメディア領域においても重要な役割を持つ。距離計算のための具体的なアプローチとして、シーケンスマッチングがある [1] [2]。シーケンスマッチングでは、2つのシーケンスを構成する各要素を対応づけ、その対応に基づきシーケンス全体の距離を計算する。この対応づけのことをマッチングと呼ぶ。マッチングにおいては、シーケンスを構成する各要素の単体としての類似性だけでなく、その順序や隣接要素の関連性、さらにシーケンス内の時間分布の違いを考慮する必要がある。

近年、最適輸送 (OT: Optimal transport) 問題 [3] [4] を用いたマッチング手法 [5] が提案されているが、本稿ではさらに隣接要素の関連性と時間分布の違いを考慮に入れた手法として Grouped OPW (G-OPW) を提案する。G-OPW では、マッチング情報を基にシーケンス内の類似要素をグ

ループ化し、グループ単位でのマッチングを行うことで、時間分布の違いを吸収したマッチングを実現する。

2. 関連研究

本稿では、OTを用いたシーケンスマッチング手法である Order-preserving Wasserstein Distance (OPW) [5] を拡張するため、その概要について説明する。OTは2つの確率分布間の距離を求める問題であるが、OPWでは2つの確率分布の各点を2つのシーケンスデータの各要素に対応させることでシーケンスデータ間の距離を定義している。2つのシーケンス $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M]$ の OPW 距離 d_{OPW} は、各要素間の距離を表す距離行列 $\mathbf{D} \in \mathbb{R}^{N \times M}$ を用いて以下で定義される。

$$d_{\text{OPW}}(\mathbf{X}, \mathbf{Y}) = \langle \mathbf{T}^*, \mathbf{D} \rangle. \quad (1)$$

ここで、輸送行列 $\mathbf{T}^* \in \mathbb{R}^{N \times M}$ は以下に定義される。

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

上式において、 $\lambda_1, \lambda_2 > 0$ は正則化パラメータ、 KL はカルバック・ライブラー情報量を示す。ここで \mathbf{T}^* の (i, j) 要素 t_{ij} は、 \mathbf{x}_i と \mathbf{y}_j 間のマッチングを表し、その値が大きいほど両者がマッチングする確率が高いと考えられる。そこで、本稿では輸送行列 \mathbf{T}^* をマッチング行列と呼ぶ。

特筆すべきは、通常の OT 問題に対して第 2 項及び第 3 項の正則化項を加えることで、順序の近い要素同士がマッ

¹ Department of Communications and Computer Engineering, School of Fundamental Science and Engineering, Waseda University, Japan

² Department of Computer Science and Communications Engineering, Graduate School of Fundamental Science and Engineering, Waseda University, Japan

a) kidsgoldppp@akane.waseda.jp

b) hiroyuki.kasai@waseda.jp

チングしやすくなり、シーケンス構造に適したマッチング行列 \mathbf{T}^* を導出している点にある。

3. 提案手法

3.1 背景課題

OPW では、正規化項によって近い順序の要素同士をマッチングしやすにしたが、実際にはあえて遠い順序同士がマッチングすべきデータ構造も存在する。それは事象の時間分布の異なるデータである。例えば「ボールを蹴る」という動作を示す2つの映像シーケンス A, B に対して、A は助走の部分が長く、B は足を振り抜いた後の部分が長く収められている場合を想定する。このように、シーケンス全体を構成する小さな事象の長さが異なることを事象の時間分布が異なるという。シーケンス A と B のように、シーケンス全体では同じことを表しながらも、その時間分布が異なるデータに対しては、同じ事象を表す要素同士 (A の助走部分と B の助走部分、など) がマッチングすることが理想的である。しかしながら、これらの要素は同じ順序には存在しない。このようなデータに対して近い順序同士をマッチングさせることは有効ではなく、全く関連のない要素がマッチングすることで、同じことを表すデータでありながら距離が大きくなってしまう。

3.2 提案手法 Grouped OPW (G-OPW)

この課題に対応するため、本稿ではシーケンス内で同じ事象を表す類似要素をグループ化し、グループ単位でマッチングすることを考える。シーケンス間で長さが異なる部分をグループ化によって一つにまとめ、グループ単位でマッチングを行うことで、同じことを表すデータに対する理想的なマッチングを近い順序同士にすることができる。このように類似要素をグループ化しグループ単位でマッチングを行う手法は存在する [6] が、本稿では、既に得られたマッチング情報からグループ化を行い、マッチングを繰り返し行うことを特に考える。

類似要素を判別するために、前節の OPW によって得られるマッチング行列 \mathbf{T} の要素値の分布のうち、水平または垂直に高い値が分布している部分に着目する。マッチング行列において、図 1 の左側に示すような水平部分がある部分は、一方のシーケンスの連続した複数要素が、もう一

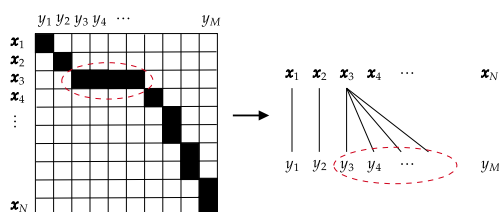


図 1 マッチング行列における水平部分と、複数要素の同一の要素へのマッチングの対応

方のシーケンスの一つの要素にマッチングしていることを表す。従って、これらの複数の要素は互いに類似していると考えられる。また、これらが連続していることから、これらの要素はシーケンス内で同じ事象を表していると考えられる。以上から、これらの要素を一つのグループと見なす。さらに、2つのシーケンスの全てのグループをそれぞれ新たなベクトルで置き換える。置換後の新たなシーケンスを用いてマッチングを再度行うことで、隣接要素の関連と時間分布の違いを考慮したグループ単位でのマッチングを実現することを考える。具体的には以下の通りである。

まずはじめに、マッチング行列 \mathbf{T} から水平または垂直なマッチング要素を探し出す。 t_{ij} は \mathbf{x}_i と \mathbf{y}_j のマッチング確率を表すことから、どの程度の大きさの値をマッチングとみなすかを定める基準が必要となる。そこで式 (2) を用いる。

$$t_{ij} \geq \alpha \cdot \max_{k \leq M} t_{ik} \quad \text{または} \quad t_{ij} \geq \alpha \cdot \max_{k \leq N} t_{kj}. \quad (2)$$

ここで、 $0 < \alpha \leq 1$ はパラメータである。すなわち、 \mathbf{T} の各行、各列について、その最大値及び最大値の α 倍以上の値を持つ要素をマッチングとみなす。 $\alpha = 1$ の場合、全ての要素について最も強いマッチングのみをマッチングとみなすことを意味する。式 (2) を満たす t_{ij} のうち、水平または垂直に $n (\geq 2)$ 個以上連続する部分を探索し、それらを一つのグループとみなす。したがって、提案手法は α と n の2つのパラメータを持ち、それらはグループ化基準の厳しさを示している。

2つのシーケンスについて全てのグループを求めた後、各グループについてグループの要素を置き換える新たなベクトルを求める。あるグループを置換するベクトル \mathbf{x}_{i^*} を式 (3) により求める。すなわち、 \mathbf{x}_{i^*} を、当該グループに対する平均ベクトルとして定義する。

$$\mathbf{x}_{i^*} = \frac{\sum_{i \in I} \mathbf{x}_i}{N_I} \quad (3)$$

ここで、 I , N_I はそれぞれ当該グループに属する要素のインデックスの集合と要素数を表す。この処理を全てのグループについて行い新たなシーケンス \mathbf{X}' , \mathbf{Y}' を得る。さらに、 \mathbf{X}' , \mathbf{Y}' について OPW によるマッチング及びグループ化を繰り返し、シーケンスがグループ化されなくなった時点で終了する。終了時点での \mathbf{X}' , \mathbf{Y}' に対する OPW 距離を提案手法による距離とする。以上の手順を Algorithm 1 にまとめる。

4. 数値実験

4.1 実験条件

提案手法について、 k 近傍法によるシーケンス分類の Accuracy (Acc) 及び MAP により評価を行う。ここで、Accuracy は分類の正解率、MAP は平均適合率の平均であ

Algorithm 1 Grouped OPW (G-OPW)

Require: 入力シーケンス: \mathbf{X}, \mathbf{Y} , パラメータ: α, n
Ensure: G-OPW 距離 $d_{G-OPW}(\mathbf{X}, \mathbf{Y})$

- 1: **while** $i < \maxiter$ **do**
- 2: OPW によるマッチング行列 \mathbf{T} を得る
- 3: 式 (2) により, マッチングの判定を行う
- 4: 水平 (垂直) に n 以上連続したマッチングを探索する
- 5: グループがなければ終了する
- 6: 式 (3) により, 各グループを単一ベクトルに置き換え, 新たなシーケンス \mathbf{X}', \mathbf{Y}' を得る
- 7: **end while**
- 8: $d_{G-OPW}(\mathbf{X}, \mathbf{Y}) = d_{OPW}(\mathbf{X}', \mathbf{Y}')$ として距離を計算する

り, いずれも高い値ほど分類精度が優れていることを示す。また, 用いたデータセットは, Spoken Arabic Digit (SAD)[7]^{*1}, High-Quality Recordings of Australian Sign Language Signs (HAS)[7]^{*2}, 及び FacesUCR[8]^{*3} である。提案手法 G-OPW におけるパラメータは $n = 2, \alpha = 1$ とし, OPW におけるパラメータは [5] に示された値を採用した。比較手法として, DTW[9] 及び OPW を採用した。

4.2 実験結果

実験結果を表 1 に示す。なお表 1 では, 最も高い値を太字, 2 番目に高い値を下線で示している。SAD データセットにおいては, 提案手法は 1 近傍法における Acc 及び MAP において最も高い値を記録した。HAS データセットにおいては, 全ての項目で DTW に次ぐ 2 番目の値を記録した。この 2 つのデータセットにおいては, G-OPW は基となる OPW による分類よりも精度を向上させることができている。一方で, FacesUCR データセットにおいては G-OPW による分類精度は OPW よりも悪化した。

5. まとめ

本稿ではシーケンスデータ間の距離計算のための新たなシーケンスマッチング手法を提案した。提案手法では, マッチング情報から類似要素を判別しグループ化しグループ単位でマッチングを行うことで, 隣接要素間の関連とシーケンス内の時間分布の違いを考慮したマッチングを実現した。提案手法から計算される距離によるデータの分類は, いくつかのデータセットにおいて既存手法による分類と比べて高い精度を記録し, その有効性を示したが, 一方で, 提案手法が有効に働かないデータも存在した。そのため, 今後は提案手法がどのようなデータ構造に対して有効に働くかという点に着目し研究を行なっていく。また, 提案手法ではマッチングとグループ化を独立に行っているた

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

*2 [https://archive.ics.uci.edu/ml/datasets/Australian+Sign+Language+signs+\(High+Quality\)](https://archive.ics.uci.edu/ml/datasets/Australian+Sign+Language+signs+(High+Quality))

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

表 1 実験結果

		DTW	OPW	G-OPW
SAD	1-NN+Acc	0.9636	<u>0.9641</u>	0.9686
	3-NN+Acc	0.9691	0.9627	<u>0.9659</u>
	6-NN+Acc	0.9732	0.9586	<u>0.9695</u>
	MAP	0.5658	<u>0.5979</u>	0.6218
HAS	1-NN+Acc	0.8135	0.7586	<u>0.7977</u>
	3-NN+Acc	0.7662	0.7271	<u>0.7444</u>
	6-NN+Acc	0.7556	0.6947	<u>0.7180</u>
	MAP	0.4461	0.4198	<u>0.4290</u>
FacesUCR	1-NN+Acc	0.9049	0.9434	<u>0.9332</u>
	3-NN+Acc	0.8863	0.9337	<u>0.9307</u>
	6-NN+Acc	0.8420	<u>0.9190</u>	0.9195
	MAP	0.6216	0.7324	<u>0.6993</u>

め, さらにグループ化の際にマッチング行列 \mathbf{T} しか考慮できない。そこで, マッチングとグループ化を同時に解く最適化問題を定義し, 距離行列 \mathbf{D} などより多くの情報をグループ化に用いる手法の提案も行なっていく。

参考文献

- [1] Jiaping Zhao and Laurent Itti. shapedtw: Shape dynamic time warping. *Pattern Recognition*, 74:171–184, 2018.
- [2] Eamonn J Keogh and Michael J Pazzani. Derivative dynamic time warping. In *Proceedings of the 2001 SIAM international conference on data mining*, pages 1–11. SIAM, 2001.
- [3] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
- [4] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- [5] B. Su and G. Hua. Order-preserving optimal transport for distances between sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2961–2974, 2019.
- [6] J. Qiu, X. Wang, P. Fua, and D. Tao. Matching seqlets: An unsupervised approach for locality preserving sequence matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2019.
- [7] Arthur Asuncion and David Newman. Uci machine learning repository, 2007.
- [8] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Annh Ratanamahatana, and Eamonn Keogh. The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019.
- [9] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.