

## 最適輸送に基づく地震波形データからのイベント検出とその分類

数理情報学専攻 48-226234 柳田 真輝

指導教員 長尾 大道 准教授

## 1 導入

1995年兵庫県南部地震を契機に、地震現象の解明・発生予測を目的として、防災科学技術研究所により高感度地震観測網 (Hi-net) をはじめとする空間的に稠密な地震観測網が整備された。Hi-net 稼働開始による主要な成果の一つとして、プレート境界面上での非火山性の低周波微動の発見が挙げられる。低周波微動は、スロー地震と呼ばれる通常の地震とは異なる地球内部起源の振動現象の一つであり、Obara (2002)[1] によって、西南日本の Hi-net 観測点群で得られた地震波形データの観測点間の空間相関解析から発見された。

低周波微動は通常の地震と関連した現象であると考えられるが、その関連性については研究途上である。低周波微動の時空間特性を詳細に調べることで、地震との関連性や、ひいてはプレート境界型巨大地震の発生予測の解明につながる可能性を秘めている。しかし、低周波微動は単なるノイズとの区別が難しいことや、波形に関する特性について未知である部分が多いことから、低周波微動検出は地震波検出に比べて困難であり、低周波微動検出のためのアルゴリズムは地震・地震波検出のアルゴリズムと同様には確立されているとは言えない。

そこで本研究では与えられた地震波形データについて、地震・低周波微動・ノイズのいずれを含むかの分類を行うアルゴリズムの構築を目的とする。その中では複数の地震波形データの時刻どうしでの対応関係について考慮するため、最適輸送距離を用いたアルゴリズム構築を行う。

## 2 準備

ベクトル  $\mathbf{a} := (a_1, \dots, a_n)^\top \in \mathbb{R}_+^n$ ,  $\mathbf{b} := (b_1, \dots, b_m)^\top \in \mathbb{R}_+^m$  及び行列  $\mathbf{C} \in \mathbb{R}^{n \times m}$  に対し、以下のように定義される最適化問題を最適輸送問題  $\text{OT}(\mathbf{a}, \mathbf{b}, \mathbf{C})$  とする。また、式 (1) の最適値を最適輸送距離とする (Kantorovich, (1942)[2])。

$$\min_{\mathbf{P} \in \mathcal{U}} \sum_{i=1}^n \sum_{j=1}^m C_{ij} P_{ij} \quad (1)$$

$$\text{s.t. } \mathcal{U} := \{\mathbf{P} \in \mathbb{R}_+^{n \times m} \mid \mathbf{P}\mathbf{1}_m = \mathbf{a}, \mathbf{P}^\top \mathbf{1}_n = \mathbf{b}\} \quad (2)$$

本研究では最適輸送距離の計算の高速化を目的として最適輸送距離を発展させたエントロピー正則化付き最適輸送距離 (Cuturi 2013[3]) と Slice-Wasserstein 距離 (Rabin *et al.*(2012)[4]) を用いる。

## 3 提案手法

地震波形データの分類のための時系列間の距離の定量化・時点どうしの対応の定量化を行うために Zhang *et al.*(2020)[5] が提案した時系列間の距離・対応を計算するアルゴリズム 1 で計算される Time-Adaptive Optimal Transport (TAOT) を用いる。5行目の  $(i-j)^2$  は時点どうしの対応に関するコストであり、時間的に矛盾のない対応が出力されにくいように設計してある。また、 $\omega$  は時間に関するハイパーパラメータである。8行目の  $\text{OT}_\varepsilon(\mathbf{a}, \mathbf{b}, \mathbf{C})$  は正則化項の係数が  $\varepsilon$  であるエントロピー正則化付き最適輸送問題を表す。分類を行うためのアルゴリズムについては、時系列間の距離に TAOT を用いた k-nearest neighbor 法を用いる。

**Algorithm 1** Time-Adaptive Optimal Transport**Input:**  $\mathbf{x} = (x_1, \dots, x_n)$ ,  $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^{d \times n}$ **Output:**  $\mathbf{x}, \mathbf{y}$  の距離と各時刻の対応関係

- 1:  $(\mathbf{x}, \mathbf{y})$  の各成分を最大値が 1 になるよう正規化
- 2:  $\mathbf{C} := (n \times n \text{ 行列})$
- 3: **for**  $i = 1, \dots, n$  :
- 4:     **for**  $j = 1, \dots, n$  :
- 5:          $C_{ij} := \|x_i - y_j\|^2 + \omega(i-j)^2$
- 6:  $\mathbf{a} := (1/n, \dots, 1/n) \in \mathbb{R}^n$
- 7:  $\mathbf{b} := (1/n, \dots, 1/n) \in \mathbb{R}^n$
- 8:  $\mathbf{P}^* := \text{OT}_\varepsilon(\mathbf{a}, \mathbf{b}, \mathbf{C})$  の最適解
- 9: **return**  $\sqrt{\sum_{i,j} P_{ij}^* C_{ij}}, \mathbf{P}^*$

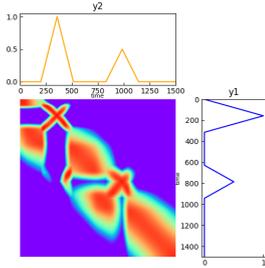
加えて、TAOT にスライス法を適用した Sliced-TAOT を提案する。このアルゴリズムは TAOT と異なり時点どうしの対応は陽に計算されないが、TAOT よりも小さい計算量で時系列間の距離が計算される。手順としてはアルゴリズム 2 の通りであり、 $K$  はサンプリング数、 $\omega$  は時間に関するコストの係数である。

**Algorithm 2** Sliced-TAOT**Input:**  $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^{d \times n}$ **Output:**  $\mathbf{x}, \mathbf{y}$  の間の距離

- 1:  $(\mathbf{x}, \mathbf{y})$  の各成分を最大値が 1 になるよう正規化
- 2: **for**  $k = 1, \dots, K$  :
- 3:  $(\boldsymbol{\varepsilon}, \varepsilon_{d+1}) \sim \text{Uni}(\mathbb{S}^d)$
- 4:  $\mathbf{x}' := (x_1^\top \boldsymbol{\varepsilon} + \omega \varepsilon_{d+1} \times 1, \dots, x_n^\top \boldsymbol{\varepsilon} + \omega \varepsilon_{d+1} n)$
- 5:  $\mathbf{y}' := (y_1^\top \boldsymbol{\varepsilon} + \omega \varepsilon_{d+1} \times 1, \dots, y_n^\top \boldsymbol{\varepsilon} + \omega \varepsilon_{d+1} n)$
- 6:  $\mathbf{x}', \mathbf{y}'$  をソート
- 7:  $s_k := \frac{1}{n} \sum_{i=1}^n (x'_i - y'_i)^2$
- 8: **return**  $\left( \frac{1}{K} \sum_{k=1}^K s_k \right)^{1/2}$

## 4 人工データに対する数値実験

地震波形の包絡線に似せた二つの三角波に TAOT を適用し、その最適解  $\mathbf{P}^*$  による時系列の時刻間での対応関係をヒートマップにより可視化する。

図 1:  $(\omega, \varepsilon) = (10^{-6}, 10^{-2})$  での  $\log P_{ij}^*$ 

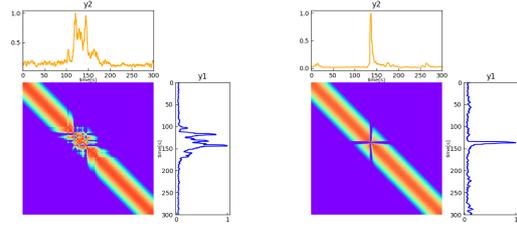
時間に関するコストの係数を  $\omega = 10^{-6}$ ,  $\text{OT}_\varepsilon(\mathbf{a}, \mathbf{b}, \mathbf{C})$  の正規化項の係数を  $\varepsilon = 10^{-2}$  としたときの結果は図 1 の通りである。ピークにあたる時刻どうしでの  $P_{ij}^*$  が大きな値をとっており、時系列のイベント時刻の対応が  $P_{ij}^*$  によって抽出されることが示唆される結果を得た。

## 5 実データに対する適用

### 5.1 TAOT の最適解の挙動

人工データと同様の設定・パラメータで、同一の地震または微動を含んだ二つの地震波形データに TAOT を適用し、その最適解  $\mathbf{P}^*$  による時系列の時刻間での対応関係をヒートマップにより可視化する。 $(\omega, \varepsilon) = (10^{-6}, 10^{-2})$  での結果は図 2a,2b の通りである。人工データと同様にピークにあたる時刻どうしでの  $P_{ij}^*$  が大きな値をとっており、時系列のイベント時刻の対応が

$P_{ij}^*$  によって抽出されることが示唆される結果を得た。



(a) 微動データ

(b) 地震データ

図 2: それぞれのイベントどうしにおける  $\log P_{ij}^*$ 

## 5.2 k-nearest neighbor 法によるイベント分類

距離指標に TAOT または Sliced-TAOT を用いた k-nearest neighbor 法によるイベント分類を行う。実験は、(1) 地震・微動・ノイズの 3 種類のイベントのいずれかを含む地震波形データを用いる。(2) それぞれのイベントについて 30 個のデータを辞書データ、10 個のデータをテストデータとして使用し、k-nearest neighbor 法によってテストデータの分類を行い、その正解率を分析する、という手順で行った。近傍の数を  $k = 5$  として実験を行った結果、TAOT/Sliced-TAOT のいずれの手法においても、6~7 割程度という比較的良好な精度で分類を行えることが分かった。

## 6 結論と今後の展望

本研究では、地震波形データからの地震・低周波微動の検出とその分類のためのアルゴリズムを提案し、人工データや実データへの適用により、その有用性が示唆された。今後の課題として、より大規模なデータセットへの適用や提案指標の理論的解析等が挙げられる。

## 参考文献

- [1] K. Obara. Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science*, Vol. 296, No. 5573, pp. 1679–1681, 2002.
- [2] L. Kantorovich. On the transfer of masses. *Doklady Akademii Nauk*, Vol. 37, No. 2, pp. 227–229, 1942.
- [3] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, Vol. 26, , 2013.
- [4] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision*, pp. 435–446. Springer, 2012.
- [5] Z. Zhang, P. Tang, and T. Corpetti. Time adaptive optimal transport: A framework of time series similarity measure. *IEEE Access*, Vol. 8, pp. 149764–149774, 2020.