

教師なし学習による連続波形記録の自動分類：階層的クラスタリングの導入

小寺 祐貴 (気象庁気象研究所)

1. はじめに

地震計の連続波形記録には、自然地震、人為的活動、機器ノイズといった様々な波形が記録されている。連続波形記録を自動的に分類することができれば、地震計周辺の地球物理的現象の理解や、緊急地震速報といったリアルタイム処理で用いる観測機器の品質管理につながると期待される。そこで我々は、様々な観測環境下に設置された地震計に適用可能となるよう、教師なし学習による連続波形記録の自動分類手法の開発を行っている。

小寺・酒井 (2020, 地震学会) では、まず特徴量としてランニングスペクトルを計算した後、周波数領域で k-means 法に基づくクラスタリングを行い、更に時間領域上での隣接関係を利用してスペクトラルクラスタリングを行う手法を提案した。しかしながら同手法では、最終的に得られるクラスタ数をスペクトラルクラスタリングのハイパーパラメータとして前もって与える必要があり、最適なハイパーパラメータをあらかじめ決めておくのは容易ではない。本研究では、階層的クラスタリングを導入することで、あらかじめクラスタ数を与えることなく、より柔軟な形で波形分類を行う手法を提案する。

2. 提案手法

提案手法は小寺・酒井 (2020) の手法と同様、次の3つのステップに分かれる：(1) 特徴抽出、(2) 周波数領域でのクラスタリング、(3) 時間領域でのクラスタリング。

- (1) 特徴抽出：ランニングスペクトルを特徴量として採用する。4秒窓でランニングスペクトルを0.1秒ごとに計算する。10段階の周波数帯域に分けたフィルタバンクに通すことで、10次元ベクトルに変換する。
- (2) 周波数領域でのクラスタリング：周波数空間上において、データ母集団から代表点を2000点選び、各データを最近傍の代表点に所属させることでクラスタリングを行う。代表点はデータ母集団からのランダムサンプリングで決定するが、データ不均衡（定常ノイズのデータ数が最多）の影響を抑えるため、データ同士の距離に基づきデータをいくつかのグループに分けた上で、それぞれのグループから代表点のサンプリングを行う工夫をする。
- (3) 時間領域でのクラスタリング：スペクトラルクラスタリングでは、グラフの隣接行列をもとにカーネル主成分分析を行った後、k-means法でクラスタを決定しているが、k-means法を行う部分をWard法による階層クラスタリングに置き換える。隣接行列はマルコフ連鎖を仮定した遷移行列から求め、主成分分析の次元数は10とする。

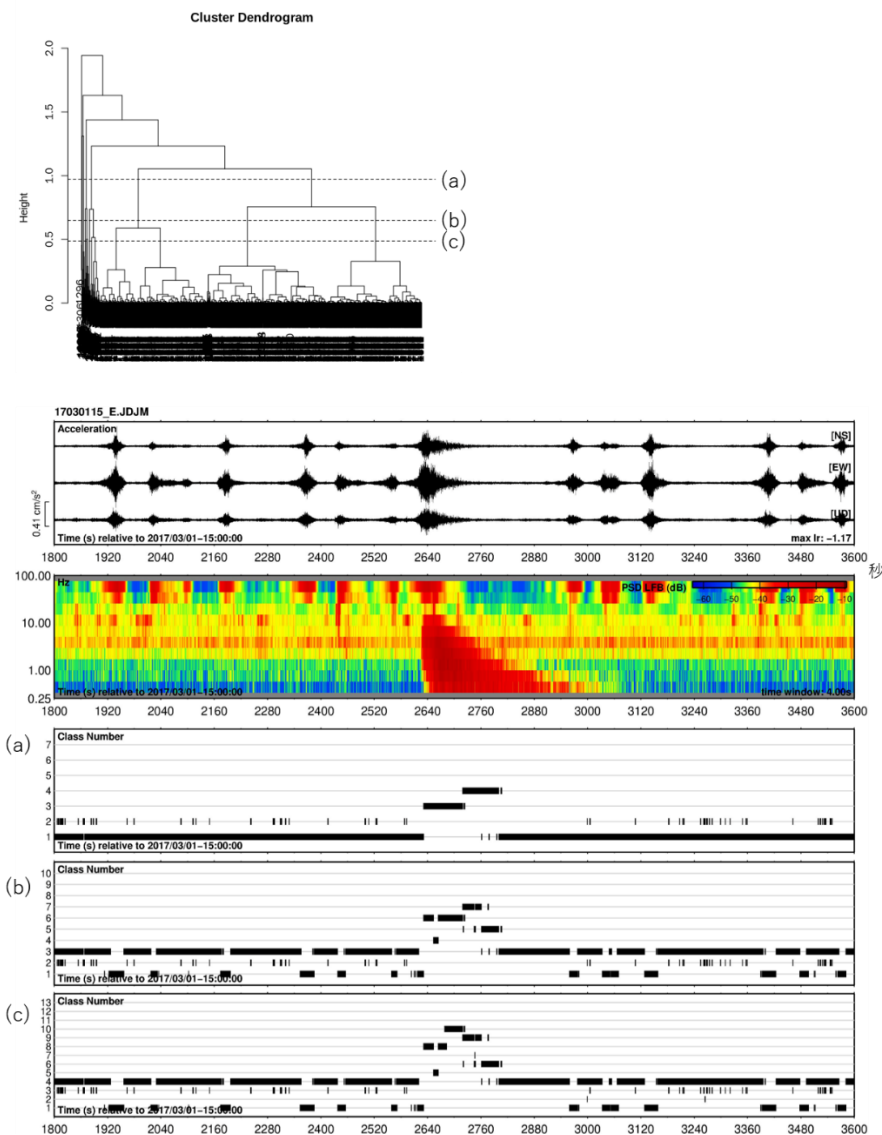
3. テストデータを用いた検証

提案手法を2017年3月1日～7日に記録されたMeSO-net観測点E.JDJM（地下鉄近傍の観測点；川北・酒井, 2009）の連続波形記録に対して適用した。階層的クラスタリング

により得られたデンドログラムと、ある時間帯におけるクラス分類結果を図に示す。デンドログラムを見ると、高い位置において小さいクラスタに分離されているグループと大きなクラスタとして分離されているグループがあり、前者は地震に関するグループ、後者がノイズに対応していると考えられる。デンドログラムをそれぞれ最大高さの 1/2 でカットした場合、定常的に混入しているノイズは高レベルと低レベルの 2 クラスに分けられていた。最大高さの 1/3 でカットした場合、高レベルのノイズは 2 つのクラスに分離され、そのうちの 1 つは鉄道ノイズに対応するクラスであった。

これらの結果は、デンドログラムをどのようにカットするのが最適かという課題は残るものの、階層的クラスタリングによって連続波形記録上のシグナルを分類・抽出できる可能性を示唆している。

謝辞：本研究には東大地震研が管理する MeSO-net の観測波形を利用しました。



図：階層的クラスタリングにより得られたデンドログラムとクラス分類結果。(a), (b), (c)はデンドログラムをそれぞれ最大高さの 1/2, 1/3, 1/4 でカットしたときの分類結果に対応する。