

An unsupervised automatic classification algorithm for continuous seismic records using a nonparametric Bayesian approach

*Yuki Kodera¹, Shin'ichi Sakai², Takahito Nishimiya¹

1. Meteorological Research Institute, Japan Meteorological Agency, 2. Earthquake Research Institute, the University of Tokyo

地震計の連続波形記録には、地震、人為的活動、機器ノイズなどの様々な要因で生じた信号が記録されている。これらの信号を自動的に抽出・分類することができれば、地震計付近で発生している定常的・非定常的な自然現象の自動検出や、リアルタイム処理に用いている地震計の自動ヘルスマニタリング等につながると期待される。本研究では、これらへの第一歩として、機械学習を用いて連続波形記録を自動的に分類する手法の開発に取り組んでいる。個々の地震計の観測環境は様々であることから、教師データがあらかじめ存在することは想定せず、教師なし学習を前提として手法を構築する。

小寺・酒井（2018, 地震学会）は、10次元に集約した離散的な周波数スペクトルを特徴量として、周波数領域および時間領域で多段階のクラスタリングを行うことで、連続波形記録を10種類の時系列データに分類する手法を提案し、自然地震、鉄道ノイズ、バックグラウンドノイズが自動的に分類可能であることを示した。しかしながら同手法は、時間領域でのクラスタリングを行う際にスペクトラルクラスタリングを用いているという問題点がある。スペクトラルクラスタリングは、分類されるクラスタ数をハイパーパラメータとして陽に与える必要があるため、その値を設定する際に恣意的な判断が入り得る。そこで本研究では、分類の客観性をより高めるため、ノンパラメトリックベイズモデルのひとつである無限関係モデル（IRM; Kemp et al., 2006）を用いて時間領域のクラスタリングを行った。IRMは無数個のクラスタの存在を許容した生成モデルであり、学習データに基づいて有効なクラスタ数が自動的に計算される。先行研究同様、時系列データがマルコフモデルで遷移すると仮定して遷移行列を生成し、その行列に対してIRMを適用した。

上記手法をMeSO-netの観測点E.JDJM（川北・酒井, 2009）および気象庁の紀伊半島沖OBSの臨時観測点TN042A（山崎・他, 2008）における72時間の連続波形記録（期間はそれぞれ2017年3月1～3日, 2004年9月22～24日）に適用した。IRMにより、E.JDJMは10種類、TN042Aは6種類のクラスに分類され、学習データに応じてクラス数が自動的に調整されることを確認した。E.JDJMにおける地震と鉄道ノイズ、TN042Aにおける振幅の大きな余震と低周波微動は別々のクラスに分類された。一方で、それぞれの信号とクラスは基本的には一対一の対応にはなっておらず、例えばE.JDJMの鉄道ノイズは5種類程度のクラスに細分化されるといったことが生じた。より直観的に理解しやすい分類とするためには、細分化されたクラスを更にまとめ上げるといった処理が必要になるだろう。