

## 深層 Hashing による類似波形をもつ地震性イベントの超省メモリ・高速探索 Detection of seismic events having similar waveforms by a deep hashing technique

○直井誠・平野史朗

○Makoto NAOI, Shiro HIRANO

We developed a deep hashing network that compresses waveform information into 64-bit binary codes and attempted event detection on 16-channel, 30-minute acoustic emission records (discretized to ~35 million windows) recorded at 10 MHz sampling during a hydraulic fracturing experiment in the laboratory. We used hash codes of 6057 template events that were cataloged using conventional automatic processing methods and searched for windows showing small Hamming distances against the templates, detecting 16,224 new events. We also tried a similar waveform search without templates by calculating Hamming distances among the 35 million windows. This corresponds to autocorrelation analysis based on cross-correlations. Although such analyses are usually difficult to apply to large datasets owing to their high computation cost, it took only 15.5 hours under 120-thread parallelization. This deep hashing approach significantly reduces calculation costs in similar waveform search problems, likely enabling the application to a large-scale dataset.

### 1. はじめに

波形相関に基づく類似波形探索は、Repeating earthquake の抽出や、ノイズに埋もれた振幅の小さいイベント波形の抽出で重要な役割を果たしているが、計算コストが大きく大規模なデータセットへの適用は簡単ではない。

この問題は、Locally sensitive hashing を用いた近似近傍探索を地震波形に適用することである程度解決できる (FAST アルゴリズム; Yoon et al. 2015). この方法では、地震波形のスペクトログラムから作成したバイナリコードを、ランダム置換を用いた Hashing でさらにコンパクトなコードに変換し、これを元に類似要素探索を実施する。この Hashing では、特徴が似た波形から同じ、あるいは似たコードが得られるので、探索対象の全ての波形ペア間での総当り探索をせずに類似要素を見つけることが可能である。例えば、連続波形の全ての切り出し窓の間の相互相関係数を計算して類似波形が記録された窓を探索する autocorrelation 解析 (Brown et al. 2008) では、 $O(N^2)$  の計算コストがかかるが、Hashing を利用した方法では、Hashing に  $O(N)$ 、類似要素探索には (理想的には)  $O(1)$  のコストしかかからない。

Yoon et al. (2015) が提案した FAST アルゴリズムは一定の成果を挙げているが、単純な Hashing を用いつつ精度を担保するため、1 つの波形に対し

て数百回の Hashing を行い、同数の Hash table を保持する必要がある。そのためメモリ消費が大きく、これが適用するデータサイズの上限を規定する。Yoon et al. (2015) の例では、1 ch, 100 Hz sampling で収録された 6 ヶ月分の波形記録の処理が限界であった。その後のチューニングにより、27 channel, 6-11 年のデータセットに対して適用するに至っているが (Yoon et al. 2019), それなりの計算機環境を必要としており、さらに大きなデータセットに対して適用することは簡単ではない。

多数の Hashing を行うことに起因するメモリ消費については、より複雑な関数を用い、1 回の Hashing で波形の情報を豊富に含むコンパクトなバイナリコードを生成できれば解決できる。画像・音声認識の分野では、このような Hash 関数作成に深層学習ネットワークが用いられるようになっており、大規模データセットの高速検索が実現されている。本研究では、深層 Hashing と呼ばれる同手法を、室内実験で得られた Acoustic Emission (AE) の連続記録に適用して、大規模データセットでの類似波形探索を試みた。

### 2. 手法

本研究では、Huang et al. (2017) が提案した、類似画像検索における Deep Hashing の手法をベースとして、地震・AE 波形記録の Hashing のための深

層学習ネットワークを設計した。彼らはネットワークの学習のため、Anker 画像とそれと同じカテゴリに属する positive sample, 及び異なるカテゴリに属する negative sample の3つの画像の組(Triplet)を多数用意した。次に、各画像から1次元実数値配列(embedding vector)を求めるための深層学習ネットワークを作成し、各 Triplet に対する出力から計算できる Triplet Loss (Weinberger & Saul 2009), Quantization Loss, Entropy Loss, と呼ばれる損失の重み付き和を最小化するように訓練した。それぞれの画像から求めた embedding vector の各成分を2値化したものが、Hash コードとなる。本研究では、彼らの提案したネットワークを1次元波形に対応するように修正し、Triplet Loss を Improved Triplet Loss (Cheng et al. 2016)に置き換えたものを使用した。得られる embedding vector の配列サイズは1x64とし、最終的に64bitのHashコードが得られるモデルを作成した。

このネットワークを、室内水圧破碎実験(Tanaka et al. 2021)中に取得したAEデータを用いて訓練した。データは24ch, 10 MHz samplingで連続収録されており、そのうち感度の良い広帯域センサ16個で計測されたデータを用いる。訓練にはTanaka et al.(2021)が古典的な自動処理手法の組み合わせで作成したAEカタログにリストされている6057イベントの波形記録を用いた。

### 3. 結果

上記の6057イベントをテンプレートとして、連続波形記録中の似た波形を持つイベントを探索する、Hashコードベースのテンプレートマッチングを実施した。まず、得られた深層 Hashing モデルを、上記の6057イベントの波形記録に適用することでこれらのHashコードを取得した。次に30分の連続波形記録を、それぞれが50%重複する1024 sample 長の窓約3500万個に分割し、これらに対しても同様にHashコードを取得した。得られた6057個のテンプレートイベントと、3500万個の窓の間でハッシュコード間のHamming距離(値が異なるbitの数)を計算し、その16ch平均値を求めた。得られた平均Hamming距離の分布において、平均値から標準偏差の6倍以上離れているという基準をもちいて、小さなHamming距離を持つ窓を抽出し、さらに重複検出とChannel間の走時の整

合性が保たれていないものを除去する処理を行った所、新たに16,224個のイベントを検出できた。

上記3500万個の窓における16chのハッシュコードの合計サイズは4.5GB程度で、2023年現在の一般的なPCにおいて全てを一括してメモリに置くことは容易であり、ディスクへのInput/Outputなしに、Autocorrelation問題に相当する総当り計算を実施することも可能である。その実現可能性を検討するため、実際にHamming距離の総当り計算を実施したところ、120スレッド並列化のもとで、わずか15.5時間で計算を終了させることができた。

波形をコンパクトなバイナリコードに変換する本手法は、計算量の大きなテンプレートマッチング問題や、auto correlation問題を、近年増加を続けている大量の地表地震観測連続データを始めとした大規模なデータセットに対して適用する助けになると期待できる。

### 参考文献

- Brown, J.R. et al. 2008. An autocorrelation method to detect low frequency earthquakes within tremor, *Geophys. Res. Lett.*, 35, L16305.
- Cheng, D. et al. 2016. Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1335–1344.
- Huang, S. et al. 2017. Unsupervised Triplet Hashing for Fast Image Retrieval, *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, 84–92
- Tanaka, R. et al. 2021. Preparatory acoustic emission activity of hydraulic fracture in granite with various viscous fluids revealed by deep learning technique, *Geophys. J. Int.*, 226, 493–510.
- Weinberger K.Q. & Saul, L.K., 2009. Distance metric learning for large margin nearest neighbor classification. *JMLR*, 10, 207–244.
- Yoon, C.E. et al. 2015. Earthquake detection through computationally efficient similarity search., *Sci. Adv.*, 1, e1501057.
- Yoon, C.E. et al., 2019. Unsupervised Large-Scale Search for Similar Earthquake Signals, *Bull. seism. Soc. Am.*, 109, 1451–1468.