

教師なし学習に基づいた突発性雑音分類システムの開発 -大型低温重力波望遠鏡 KAGRA データへの適用に向けて-

上島元[†] 苔山圭以子[‡] 小坂井千紘[§] 中平勝子[†] 押野翔一[‡] 鹿野豊[¶]
高橋弘毅^{||} 内山隆[‡] 鷲見貴生[§] 山本尚弘[‡] 横澤孝章[‡]
長岡技術科学大学[†] 東京大学[‡] 国立天文台[§] 慶應義塾大学[¶] 東京都市大学^{||}

1 はじめに

大型低温重力波望遠鏡 KAGRA は、時空の歪みが光速で伝搬する現象である重力波を観測するための装置である。重力波は極めて微弱な信号であるため、KAGRA の開発ではあらゆる雑音の原因または影響を排除していくことが重要である。特に懸念されるのは、地面振動や雷などの環境的な外乱により発生する突発性雑音である。突発性雑音は、原因毎に特徴的な時間-周波数特性を有すると考えられており、近年では、Gravity Spy [1][2] 等の研究により、教師あり学習に基づいた分類手法が適用されつつある。しかし、突発性雑音が原因毎に発生頻度が異なり、真の原因数は未知であることが、ラベル付け問題を困難にしているため、客観的な分類ができていない。以上を踏まえ、本稿では、教師なし学習に基づいた突発性雑音の分類システムを提案する。また、Bahaadini ら [2] による突発性雑音データに適用し、i) 類似性に基づくデータ検索、ii) クラスタリング、への利用の観点から評価をする。

2 提案手法

提案手法では、以下の手順により突発性雑音データを分類する。

1. Variational Autoencoder [3] を用いて、データ空間 $\mathcal{X} \subset \mathbb{R}^{D_x}$ を低次元の潜在空間 $\mathcal{Z} \subset \mathbb{R}^{D_z}$ ($D_z < D_x$) へ射影する。
2. Invariant Information Clustering [4] を用いて、潜在空間を M 通りでクラスタリングする弱分類器を得る (クラスタ数は C)。
3. 弱分類器の出力を連結したハイパーグラフ ($C \times M$ 次元) を、主成分分析により次元削減し、特徴量空間 $\mathcal{W} \subset \mathbb{R}^{D_w}$ を得る。
4. 特徴量空間 \mathcal{W} に基づいて、コサイン類似度行列を生成する。
5. コサイン類似度行列から、Spectral Clustering により、最終的なクラスタリング結果を得る (クラスタ数は K)。

3 実験

本実験では、Bahaadini ら [2] による 22 種類の突発性雑音クラスと 8583 個の時間-周波数マップ画像を用いる。この画像は、突発雑音の発生時刻を中心に $\pm 0.25, 0.5, 1.0, 2.0$ 秒の異なる時間幅で構成される時間-周波数マップ (解像度は全て 224×224) である。また、各画像には一つの突発性雑音クラスが対応している。本実験に用いたパラメータを以下に示す。

$$D_x = 4 \times 224 \times 224, \quad D_z = 512, \quad M = 10 \\ C = 100, \quad D_w = 64, \quad K = 50$$

Development of Unsupervised-Learning Based Classification System for Transient Noise - Towards an Application to Data from Large-scale Cryogenic Gravitational Wave Telescope KAGRA-

[†] G. Ueshima, K. Nakahira (Nagaoka University of Technology)

[‡] K. Kokeyama, S. Oshino, T. Uchiyama, T. Yamamoto, T. Yokozawa (Tokyo University)

[§] C. Kozakai, T. Washimi (National Astronomical Observatory of Japan)

[¶] Y. Shikano (Keio University)

^{||} H. Takahashi (Tokyo City University)

3.1 結果と考察

3.1.1 類似性に基づくデータ検索への利用

抽出した全データ間のコサイン類似度に基づいて、類似性が高いデータの検索を行った。図1に、無作為に抽出した突発雑音のデータ（3例）と検索された類似データを示す。

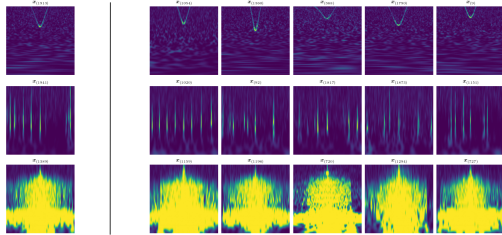


図1 無作為抽出データ（左）と類似データ（右）

図1より、特徴の類似性を捉えて検索できていることが分かる。これは、提案手法が、類似した複数のデータを用いて統計的な原因探索を行うために利用できることを示している。

3.1.2 クラスタリングへの利用

クラスタリング結果の比較のために、先行研究によるラベルと、提案手法による予測ラベルの混合行列を求めた（図2）。

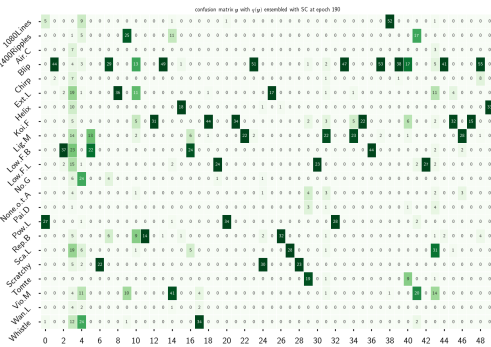


図2 先行研究によるラベルと提案手法による予測ラベルの混合行列

図2より、提案手法では、先行研究において共通した特徴を持つクラスのデータが独立して分類されていることが分かる。また、先行研究における None_of_the_Above クラスや Wandering_Line クラスのような共通した特徴を持たないデータは独立して分類されていないことが分かる。さらに、先行研究において単一のクラスに分類してい

たデータを複数のクラスに分類している場合があることが分かる。これは、提案手法が、特徴量空間におけるデータのまとまりを考慮した客観的なクラスタリングを行うために利用できることを示している。

4 おわりに

本稿では、教師なし学習に基づいた突発性雑音の分類システムを提案した。さらに、先行研究の Gravity Spy で用いられたデータに適用し、i) 類似性に基づくデータ検索、ii) クラスタリング、への利用に関する実用性を示した。

今後は、より良い分類を行うため、突発性雑音データの時間-周波数マップ以外の特徴量の利用、特徴量空間の生成過程の最適化やチューニング、分類結果の解釈性の向上についての検討を進めていく。

本研究は、JSPS 科研費 17H06358,20H04731 の助成を受けたものである。

参考文献

- [1] M. Zevin, S. Coughlin, et al., “Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science”, *Classical and Quantum Gravity* 34, 064003 (2017).
- [2] S. Bahaadini, V. Noroozi, et al., “Machine learning for Gravity Spy: Glitch classification and dataset”, *Information Sciences*, vol.444, pp.172-186 (2018).
- [3] D. Kingma, M. Welling, “Auto-Encoding Variational Bayes”, *ICLR* (2013).
- [4] X. Ji, J. Henriques, A. Vedaldi, “Invariant Information Clustering for Unsupervised Image Classification and Segmentation”, *ICCV* (2019).