

位相的データ解析を用いた 非定常時系列の変動特性の可視化とクラスタリング

東京理科大学大学院工学研究科経営工学専攻 黒木 裕鷹

東京理科大学工学部情報工学科

塩濱 敬之

1 はじめに

膨大な特徴量をもつ高次元データ解析において、どのように高次元の特徴量を低次元で表現するのは重要な研究課題である。近年、位相的データ解析による高次元データ解析が盛んに提案されており、複雑な高次元データの位相的情報を抽出する手法として注目を集めている。位相的データ解析では主として、ホモロジーの発生と消滅に着目するパーシステントホモロジー [1] によりデータの位相的情報を抽出する。ここでホモロジーとは、連結成分、わっか、空洞など、「穴」に着目した代数的対象である。このパーシステントホモロジーを可視化したパーシステントダイアグラム [2] を解析することによる知識発見や、変数間のネットワーク構造の解明に役立つことが期待されている。パーシステントダイアグラムは2次元であるために情報が多く、様々な要約が提案されてきた。例えば、大きなパーシステンスを持つホモロジーに着目した Landscape [3] や、ホモロジーの数に着目した Betti sequence [4] などの統計量が提案され、金融危機の検知や非定常時系列の変化点探索などに応用されている。

本研究では、非定常時系列データの位相情報の可視化と、その特徴量を利用した教師なしクラスタリングの手法を提案する。また例として、自動車工場の生産ラインで観測されるロボットアームの振動データの解析を取り上げ、振動データの時系列構造の可視化を通して故障・交換にかかる生産ライン停止時間の短縮に貢献することが可能であることを示唆した。

2 1次元時系列データに対する位相的データ解析

データからパーシステントホモロジーを得るには点群 (point cloud) を与える必要がある。1次元の時系列データから点群を得るために、遅れ時間座標を利用した埋め込み (embedding) がしばしば用いられる。点群の各点を中心とする円を考え、円の半径 r を大きくしていった時に構成される単体の発生と消失を元にパーシステントダイアグラムを得ることができる。パーシステントダイアグラムの統計量である Betti sequence や Landscape は、半径 r の時系列であると解釈できる。そのため、時系列間の非類似度としてしばしば用いられる DTW (Dynamic Time Warping) [5] を用いることができる。本研究では時系列データの位相的解析の例として、自動車工場の生産ラインで観測されるロボットアームの振動データを扱う。減速機を交換する前と後の振動データに対して Betti sequence と Landscape を求め、それぞれの非類似度行列を DTW により作成し、交換前のデータが交換後のクラスタから分離できることを確認した。多次元尺度構成法 (Multi-Dimensional Scaling : MDS) でそれぞれの非類似度行列を1次元の空間に再構成し、Betti sequence に由来するものと Landscape に由来するものの2次元空間で可視化することにより、効果的に知見が得られることを示した。

参考文献

- [1] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," *Discrete & Computational Geometry*, vol. 28, no. 4, pp. 511-533, 2002.
- [2] H. Edelsbrunner and J. L. Harer, "Computational topology," *American Mathematical Society*, 2010.
- [3] P. Bubenik, "Statistical topological data analysis using Persistence Landscapes," *Journal of Machine Learning Research*, vol. 16, no. 1, pp. 77-102, 2015.
- [4] Y. Umeda, "Time series classification via topological data analysis", 人工知能学会論文誌., vol. 32, no. 3, pp. D-G72 1-12, 2017.
- [5] D. Berndt and J. Clifford, "Finding patterns in time series: a dynamic programming approach," *Advances in knowledge discovery and data mining*, pp. 229-248, 1996.