

トポロジカルデータアナリシスと 時系列データ解析への応用

Topological Data Analysis and Its Application to Chronological Data Analysis

● 梅田 裕平 ● 金児 純司 ● 菊地 英幸

あらまし

近年、AI(人工知能)技術の実用化が急速に進んでおり、その牽引役であるDeep Learningをはじめとする機械学習技術に注目が集まっている。しかし、機械学習技術のベースとなっている統計手法を用いたデータ解析によって、データが本来持つ情報が失われることも知られるようになってきた。こうした情報を十分に活用するために、筆者らは「データの形」が持つ情報に着目し、それらの情報を解析するトポロジカルデータアナリシス(TDA)技術を基にした機械学習技術を開発した。

本稿では、新しいデータ解析技術であるTDAと、その応用技術であるTDAを用いた時系列データ解析技術のうち、時系列Deep Learning技術、時系列異常検知技術、およびその適用事例として橋梁劣化^{きょうりょう}解析について述べる。

Abstract

The commercialization of AI (artificial intelligence) technology has accelerated in recent years, with a growing interest in various machine learning technologies such as deep learning. However, machine learning is based on statistical data analysis, and it is known today that certain information contained in such data is lost through the analytical processes. To make the most of such information, which is otherwise lost, we have developed a technology based on topological data analysis (TDA) that focuses on and analyses the “shapes of data.” This paper will explain TDA as a new data-analytical method. As applied cases of TDA, it also describes the time-series deep learning for analyzing time-series data and anomaly detection technology, with an account of a bridge deterioration assessment in which the latter was applied.

ま え が き

近年、Deep Learningの登場を契機として第三次AI（人工知能）ブームが到来し、社会への適用が進んでいる。AI技術の進展を支えているのは、ビッグデータの収集を可能にしたIoT技術の進歩と、Deep Learningをはじめとする機械学習技術である。

機械学習を含むデータ解析技術は統計的な解析手法をベースとしており、現在も活発に研究開発が行われている。一方で、統計的な解析手法はデータが正規分布などに従っていることを前提としているため、よく知られた分布に従っていない場合や適切な分布が明確ではない場合には、期待する性能が得られないことが知られている。この問題に対して、富士通研究所では統計的な手法を用いないデータ解析手法として「データの形」を捉えることで詳細な情報を把握する、トポロジカルデータアナリシス（TDA）技術の研究を行っている。

センサーなどから取得された時系列データは、取得するタイミングによって平均や分散などの統計量、周波数が大きく変化する「カオス性」と呼ばれる特徴を持つ。このような時系列データの解析に適用可能な技術開発を進め、TDAとDeep Learning技術、および異常検知技術を組み合わせる技術を開発した。これにより、センサーを用いた高度なAIサービスを生み出すことが可能になる。

本稿では、TDAとこれを用いた時系列データ解析技術、および適用事例として橋梁の劣化解析について紹介する。

統計的なデータ解析の課題

機械学習をはじめとしたデータ解析において、従来使われている統計的な解析手法では、データがある種の分布に従っているということが前提となっていた。例えば、テストの平均点や偏差値は、点数の分布が正規分布に従っていることを前提としている。

一方で、実際のテストでは高得点層と低得点層が二極化し、二つの山を持つ分布になることがある。このように、データがどの分布に従うかわからない場合や、確率分布では捉えられない情報が含まれている場合も多く見られる。近年、ビッグ

データの収集が可能になったことによって、そのような状況が頻繁に見られるようになり、従来のような統計的な解析手法だけでは十分な性能が発揮できないことがあった。

この問題に対して、従来はデータ特有の特徴を利用した個別の解析手法を構築するか、複数の分布を組み合わせて統計的に解析する際に、可能な限り情報を落とさないように拡張する方法⁽¹⁾などが採られてきた。

しかし、それらの方法に関してもデータ特有の特徴などを得るために、データに関する詳細な知識が必要となる。また、局所的にはある分布に従っているという前提が必要となるため、情報の損失が避けられないという問題があった。

TDAにおける研究開発技術

TDAは、従来の統計的なデータ解析では失われていた重要な情報を、「データの形」を把握することによって詳細に捉える技術として研究が進められている。本章では、TDAで主に研究開発されている、MapperとPersistent homologyという二つの技術について紹介する。

● Mapper

Mapperは、データの特徴となる部分を把握しやすいグラフとして表示する技術である。データを把握する上で重要な部分をまとめて一つのノードとし、連続したデータのあるノード間を線（エッジ）でつなぐことで、データの集合をグラフに変換している。Mapperはデータの集まりを2次元のグラフとして出力するため、データの分布を可視化できる。例として、三つのウイルス発生源がある場合のウイルスの分布を考える。

図-1 (a) はウイルスの分布の例である。発生源が一つの場合、ウイルスは発生源の周りで密度が高く、発生源から離れるほど密度は低くなる。しかし、複数の発生源がある場合は、分布図からどこに発生源があるのかを把握することは困難になる。ここで、データの密度を基準として、周囲より密度が高かったり低かったりする重要な箇所をノードにまとめ、グラフ化したものが図-1 (b) である。

図-1 (b) を見ると、矢印で示した三つのノードが、それとつながったノードより密度が高いことが分

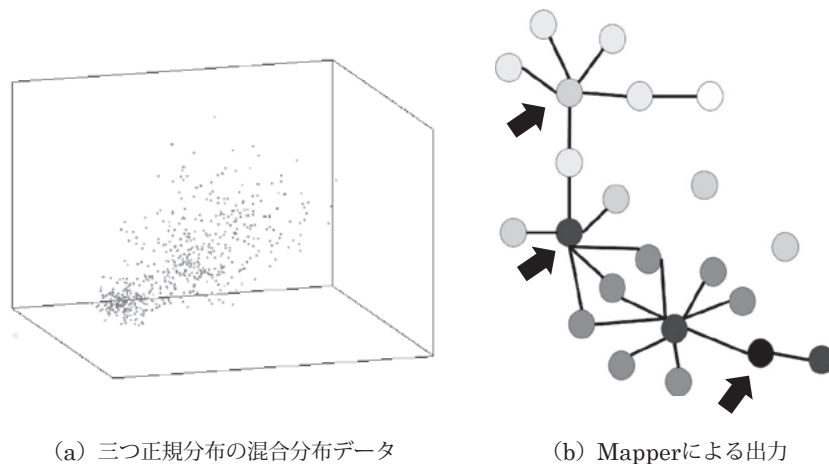


図-1 Mapperによる混合分布のグラフ化

かる。これによって、図-1(a)では認識が困難であった三つのウイルス発生源を持つということが事前の知識なしに把握できる。

このように、Mapperはそれだけでは内容の把握が困難なデータに対して、データの形状を把握しやすくできる技術である。ビッグデータに紛れて抽出が難しくなった特徴を捉えられるため、海外では金融分野におけるアンチマネーロンダリングや、ヘルスケアの分野における異なる症状の関連性の発見などで応用が始まっている。⁽²⁾

● Persistent homology

データの重要な部分を視覚的に把握するMapperに対し、データ形状の詳細を数値として捉える技術がPersistent homology⁽³⁾である。データ解析では、データの配置を把握することが重要になるが、平均や分散などの従来の統計量だけではその特徴を捉えきれない場合がある。

例えば、図-2のようにデータの配置が異なっても平均や分散が同じ場合、分布の中心に穴があるかどうかは把握できない。一方で、Persistent homologyは異なるデータの配置であれば結果が異なり、中央に穴があるなどの情報も明確な特徴として現れることが知られている。そのため、従来の統計量を用いた場合よりも詳細にデータの特徴を把握できる。

Persistent homologyの概要を図-3 (a) に示す。この技術では、データの各点を中心とした円(3次元データでは球)を考える。この円を膨らませると、円同士が結合して新たな図形ができるが、こ



図-2 同じ平均・分散を持つデータの例

の図形に含まれる穴の数の変化を見ることでデータの特徴を捉える。図-3 (b) の上段は、横軸を円の半径、縦軸を連結成分(数学的には0次の穴と呼ばれる)の数をプロットしたグラフである。図-3(a)のそれぞれの地点における円の半径と、穴の数は図-3 (b) の縦線を引いた部分に対応している。また下段は、縦軸を一般の穴の数(数学的には1次の穴と呼ばれる)をプロットしたグラフである。このグラフが入力データの配置の特徴を現すものと考えられる。

富士通研究所では、Persistent homologyを用いて数値化した情報と機械学習を組み合わせることで、高度なデータ解析を可能にする技術を開発した。時系列データの変換技術に関しても、Persistent homologyを利用している。

TDAを用いた時系列の機械学習

本章では、TDAを用いた機械学習技術を二つ紹介する。これらは「教師あり学習」「教師なし学習」として知られ、機械学習の分野では主要な学習の

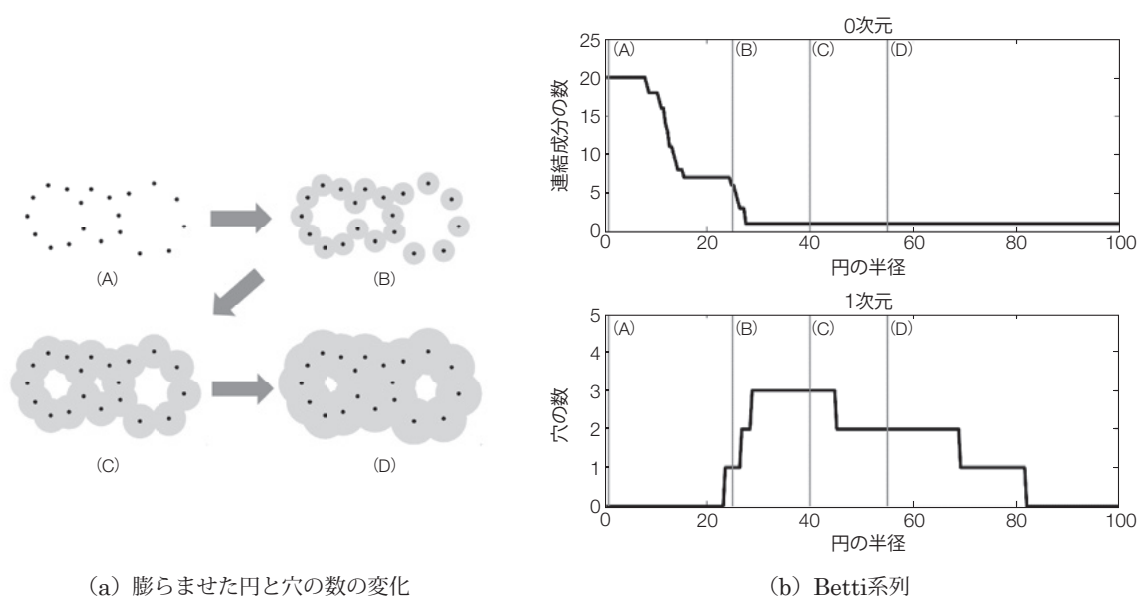


図-3 Persistent homology

枠組みを提供する重要な技術である。

● TDAと時系列データの機械学習

現在のAI技術の根幹となっている機械学習は、事前にユーザーが適切な形にデータを整形する必要があった。富士通研究所では、TDAを用いてデータを整形し、機械学習やDeep Learningを組み合わせることによって、より幅広いデータに対して高度な解析を行えるAI技術の開発に取り組んでいる。

近年のIoT技術の発達により、センサーなどから様々なデータが高性能に取得できるようになってきた。従来、センサーなどから得られた時系列データの学習には、平均や分散などの統計量に加えて周波数解析が用いられてきた。しかし、変動の激しい時系列データに対しては、このような手法では十分な性能が出せない場面も増えてきている。その原因の一つとして考えられるのが、時系列の発生に規則はあるものの、平均や分散などの統計量や周波数が一定にならない「カオス性」を持つものが存在することである。

カオス性を持つ時系列データを高い精度で学習する手法として開発した、TDAを用いた時系列データ解析の概要を図-4に示す。以下に、各ステップの概要を述べる。

(1) アトラクタ変換

TDAを用いた時系列データ解析では、カオス性を持つ時系列データであっても、変化しない支配

方程式を特徴量として学習する。支配方程式は時系列データの変動を表現する微分方程式であり、時系列データの発生源の状態が同じであれば同一であると考えられる。このため、従来の手法よりも高い性能で解析できると考えられる。

そこで、カオス理論において支配方程式を図形化する手法として知られるアトラクタ変換⁽⁴⁾を用いることで、支配方程式の情報を抽出した。アトラクタは微分方程式の解の集合であるため、方程式の解析とアトラクタの解析は同等の意味を持つ。なお、現実的に使用する時系列データは有限の長さとなるため、実際に構成されるアトラクタは解の集合を代表する有限個の点の集合となる。

(2) Persistent homologyによるベクトル化

機械学習の多くの手法では、入力となる全てのデータが同じ長さのベクトルであることが前提となっている。一方で、時系列データから構成されたアトラクタは点の集合となっているため、単純に機械学習の入力に使用することは難しい。そこで、点の配置情報を保ったままベクトル化する方法としてPersistent homologyを使用した。Persistent homologyは点の配置情報を数値化するため、ベクトル化が容易である。このようにして作成したベクトルをBetti系列と呼び、機械学習の入力として使用する。

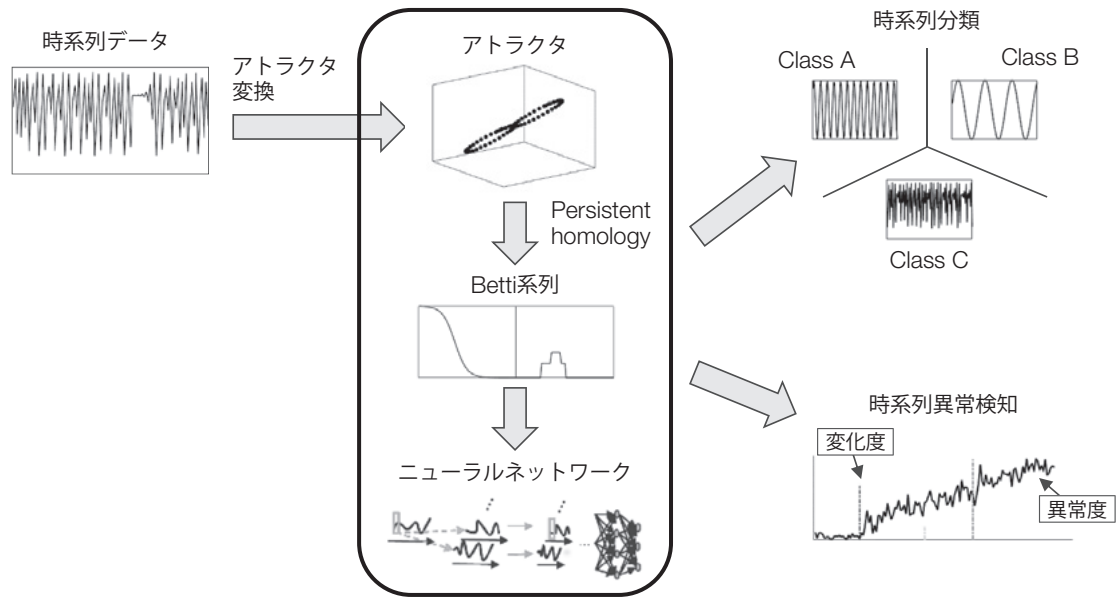


図-4 TDAを用いた時系列データ解析技術

(3) 1次元CNN (Convolutional Neural Network) による教師あり学習

アトラクタとPersistent homologyによって作成されたBetti系列は、同じ支配方程式を持つ時系列データであっても、特徴となる部分が隣接するベクトルの要素にずれて表れることがある。この特徴は、画像認識でも特徴となる物体の位置がずれることがあることと同様である。したがって、画像認識で高い効果のあったCNNはBetti系列においても効果が期待されるため、Betti系列用のCNNを構築した。画像データは2次元であるのに対して、Betti系列は1次元のベクトルである。更に、複数の独立したデータを組み合わせることがあるため、複数の1次元CNNを組み合わせさせたネットワークを独自に構成した。⁽⁵⁾

● 時系列Deep Learningの評価検証

時系列Deep Learningを評価するために、以下の三つのデータセットに対し、従来手法との比較実験を行った。データは、以下の三つを使用した。

- (1) 19種類の動作をしたときの両手・両足・胸に着けたジャイロセンサーのデータ (Gyro)
- (2) 目の開閉時の脳波データ (EEG)
- (3) 10種類の動作をしたときの筋電波形のデータ (EMG)

これら三つのデータを使用し、従来の統計処理で得られる特徴量とカオス理論の特徴量をそれぞ

れSVM (Support Vector Machine) で学習させたものを比較した。いずれのデータに対しても、従来手法と比較して20%以上正答率が向上し、時系列データの学習手法として有効であることが示された。

● TDAを用いた時系列異常部位検知技術

時系列Deep Learningは分類が明確であり、分類に応じたラベルが付けられた教師データがある場合に有効である。一方で、実際に時系列解析を行う場合、分類の基準が明確ではなく、正常状態から異常状態への変化がいつ起こったか、その予兆がいつあったかなどを検知することが目的となることも多い。そのため、TDAを用いた異常検知技術を開発した。

この技術は、時系列データから一定期間ごとにデータを取得し、それぞれのデータを時系列Deep Learningと同様にBetti系列に変換し、基準データと比較する。基準データは、異常検知の場合は正常状態として事前に取得したデータを用い、変化検知の場合は直前のデータを用いる。基準として差を計算し、差が大きくなった箇所を検出時点とする。基準と取得データの差の計算として、Deep Learningを基にして構築したものなど、複数準備している。これにより、時系列の発生源の規則変化を捉えることが可能になるため、従来手法よりも根本的な原因による異常検知・変化検知が可能

になる。

● TDAを用いた異常検知技術の実適用

富士通研究所では、TDAを用いた時系列異常検知技術を橋梁の劣化解析に適用した。以下に本事例を示す。

日本では、高度経済成長期に建設された多くの橋梁で老朽化が進んでいる。このため、それらを維持管理するための業務が急増し、メンテナンスコストの増大、技術者不足などが社会問題となりつつある。橋梁などの社会インフラの維持管理業務にICTを適用することで、これらの社会的課題の解決が期待されるようになってきている。このような状況の中、富士通はモニタリングシステム技術研究組合（RAIMS）に参画し、社会インフラの維持管理に関するモニタリングデータの蓄積・分析を担っている。

現状の橋梁点検業務は、熟練技術者の近接目視や打音検査によって損傷の検出・評価が行われている。しかし、目視のみでは構造物表面に現れた変状しか捉えることができないため、内部の損傷度合いに関する情報が把握できないという問題があった。

そのため、点検業務のICT活用による高度化に向けて、橋梁の基本部材である床版（橋の上を通る車両の重みを橋桁や橋脚に伝えるための構造物）の表面にセンサーを取り付け、振動データを収集し解析することで、損傷程度の評価を行う取り組みがなされている。しかし、これまでに振動データを活用し、スペクトル解析などの従来の手法による床版内部の損傷度合いの評価が試みられているが、点検業務の高度化にはつながっていない。

そこで富士通では、橋梁床版の内部損傷を検出・評価するために、床版表面に設置した加速度センサーからのモニタリングデータにTDAを用いた異常検知技術を適用した。⁶⁾なお、これらのデータは、RAIMSが実施した2015年度の輪荷重走行実験によるものである。

図-5に、異常検知技術で計算した異常度、変化検知技術で計算した変化度、および劣化進行の原因となる内部損傷データ（鉄筋およびコンクリートのひずみ）を示す。実験では、実際の道路橋梁には設置できない内部ひずみセンサーによるモニタリングも一部の箇所で開催しており、実際の橋梁の損傷度合いとの比較が可能となっている。

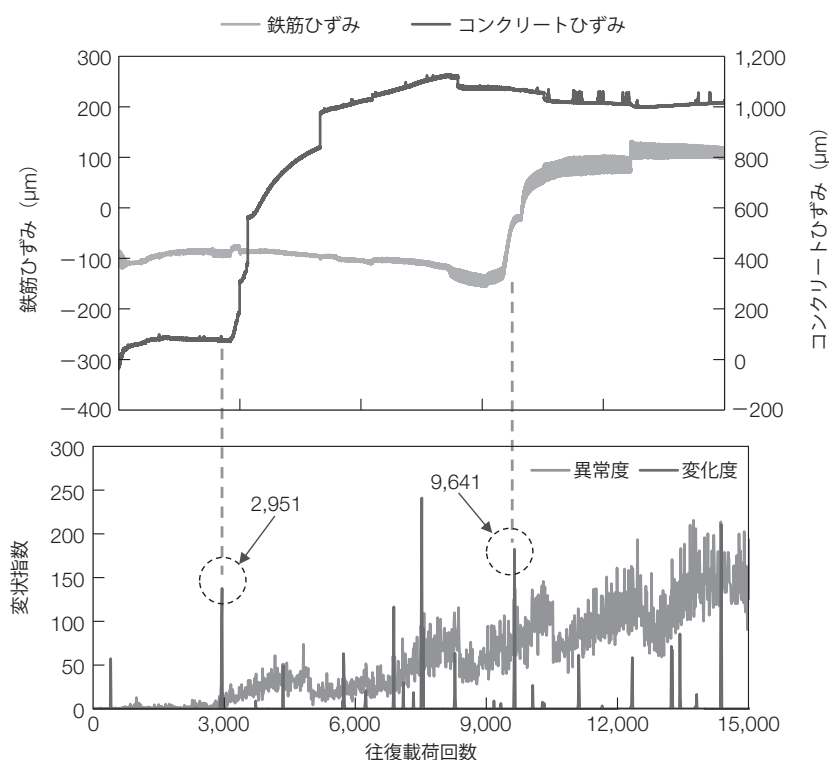


図-5 橋梁劣化解析の例

検証の結果、図-5において変化度の大きいときが内部損傷の発生を表していることが分かる。また、内部ひずみセンサーの値に変化が生じていないときに変化度が高くなっているが、これは内部ひずみセンサーの付いていない地点が損傷した影響を受けたものと考えられている。この結果は、橋梁床版の外部に設置される加速度センサーから得られるデータをTDAを用いた異常検知技術に適用することで、初期段階で内部損傷を検出できることを示している。

本事例は、TDAを用いた異常検知技術の有効性を示す一例であり、今後は本技術を実際の橋梁に適用を進めるとともに、橋梁以外の社会インフラの劣化検出へと展開していく。更に、機械システムなどの異常検知への適用も進めていく。

む す び

本稿では、データ解析の新しい技術であるTDAについて紹介するとともに、富士通研究所で開発したTDAを用いた時系列Deep Learning技術および時系列異常検知技術について解説した。

今後は、フランス国立情報学自動制御研究所 (INRIA : Institut National de Recherche en Informatique et en Automatique) との共同研究を通じて、更なる時系列データ解析の精度向上や詳細な解析手法の開発に加え、TDAの時系列解析以外への展開を図っていく予定である。

参考文献

- (1) R. Fujimaki et al. : Factorized Asymptotic Bayesian Inference for Mixture Modeling. Proceedings of the fifteenth international conference on Artificial Intelligence and Statistics (AISTATS), 2012.
- (2) Ayasdi : Improving Denial Management Using Machine Intelligence. White paper, 2017.
- (3) 平岡裕章 : タンパク質構造とトポロジー—パーシステントホモロジー. 共立出版, 2013.
- (4) 合原一幸 : カオスの数理と技術—カオス, そしてフラクタル, 複雑系への序章. 共立出版, 2013.
- (5) Y. Umeda : Time Series Classification via Topological Data Analysis. 人工知能学会論文誌, 32 (3), p.D-G72_1-12.
- (6) 金児純司ほか : 輪荷重走行試験によるRC床版の疲

劣化に関するモニタリング技術の検討 (その6) 各種分析方法とモニタリングデータによる疲労劣化の評価. 土木学会第72回年次学術講演会, 2017.

著者紹介



梅田 裕平 (うめだ ゆうへい)

(株) 富士通研究所
人工知能研究所
機械学習およびデータ解析の研究開発に従事。



金児 純司 (かねこ じゅんじ)

(株) 富士通研究所
人工知能研究所
兼 富士通 (株) 社会インフラビジネスグループ第四システム事業本部
数理解析技術とその応用の研究開発に従事。



菊地 英幸 (きくち ひでゆき)

(株) 富士通研究所
デジタル共創プロジェクト
兼 富士通 (株) 社会インフラビジネスグループ第四システム事業本部
i-Constructionの研究開発に従事。