

戦略的創造研究推進事業 CREST
研究領域「現代の数理科学と連携するモデリング手
法の構築」
研究課題「ソフトマター記述言語の創造に向けた位
相的データ解析理論の構築」

研究終了報告書

研究期間 2015年10月～2022年3月

研究代表者：平岡 裕章
(京都大学高等研究院 教授)

§ 1 研究実施の概要

(1) 実施概要

本研究課題では、ソフトマター記述言語の創造に向けた位相的データ解析理論の構築という目標に向けて、TDA 班, 表現論班, 確率論班, 統計班, MD 班の 5 つの班体制で研究を実施した. 以下に各班の代表的な成果を中心に, その研究活動の概要をまとめる.

TDA 班では, パーシステント図の逆問題理論および線形機械学習モデルの開発に関する数学研究を行った. 逆問題解法では最小生成元探索法と体積最適化法の 2 種類を提案し, いずれの方法とも材料科学をはじめとする応用研究において重要な役割を果たした. 特に線形機械学習と融合させることで, 学習結果に幾何学的な意味を与えることが可能となり, データの背後にある現象の理解につながる「interpretable AI」の一例として, 世界的にも注目されている. 応用研究では, これらの開発手法をガラス, 粉体, 高分子などのソフトマターの構造解析に適用し多くの成果を挙げた. またパーシステントホモロジー計算ソフトウェア「HomCloud」を開発し, これをもとに複数の企業との共同研究に発展させた. さらには研究代表者班として, その他の班との共同研究も多く実施し, チーム内の連携体制の強化に努めた.

表現論班では, パーシステントホモロジーのマルチパラメータ化を可能とする, 一般化パーシステント加群の分解論に関する数学研究に取り組んだ. その動機としては, MD 班が扱うタンパク質のフォールディング現象や TDA 班が取り組む材料科学の問題において, 空間および時間の両方の意味で階層的にデータ解析をする必要があり, その数学理論を構築することは位相的データ解析の分野で最重要課題となっている. そこで表現論班では, パーシステント加群を可換格子クイバーの表現の形で表示し, bocs 理論を用いてこの加群を直既約分解する研究を実施した. また研究期間後半には, マルチパラメータの設定における区間分解論および区間近似理論を提案した. これらの概念は, まだプレプリントの状態ではあるが世界的にも注目を集めており, パーシステントホモロジーのマルチパラメータ化の研究において今後重要な発展をしていくものと思われる.

確率論班ではデータ解析を動機としたランダムトポロジー理論を開拓し, 位相的データ解析の確率論的枠組みを構築する研究を実施した. 特にランダムな単体複体やセル複体の構造解析とパーシステント図の解析, 距離空間や多様体上のランダム点過程やランダム場が定めるパーシステント図の解析を中心に研究を展開し, その過程であらわれる周辺の数学的な問題にも取り組んだ. 研究期間初期には, 単体複体の増加過程のパーシステント図解析として, エルデシュ・レンニグラフ過程の複体版である Linial-Meshulam 複体過程やランダム Clique 複体過程に対して, Frieze の $\zeta(3)$ -定理の拡張に関して成果を得た. また中期から後期にかけては, R^n 上のランダム点過程が定めるパーシステント図の解析に取り組んだ. ここでは点群データに対して, Čech 複体や Rips 複体を含む κ -複体 (κ -フィルトレーション) という概念を導入し, エルゴード的な定常点過程の κ -フィルトレーションのパーシステント図が, 無限体積極限で確率 1 である一つの測度に濃収束することを証明した.

統計班では, 位相的データ解析の統計的枠組みを確立し, 統計データ解析にパーシステントホモロジー解析を応用することで, 位相的統計理論と呼ぶべき新たな分野創出を目指す研究を実施した. パーシステントホモロジーを実際のデータ解析に適用する際には, データの有限性や観測ノイズなどによってランダム性を持つことになり, そこから有効な情報を抽出するためには, 統計的に有意な解析手法を確立することが必須である. そのために, パーシステント図に対するデータ解析手法を包括的に適用可能とする Persistence Weighted Gaussian Kernel (PWGK) 法とよぶカーネル法を開発した. これは通常のガウスクーネルに重み関数をかけたカーネルであり, 大局的なスケールとノイズに近い小さいスケールをうまく調節することによって, パーシステントランドスケープやパーシステントイメージなどの類似の既存手法に比べて優位性を持つことを示した. また, このカーネルの定めるパーシステント図間の距離が, 安定性という理論的性質を持つことを数学的に証明した. これらの一連の成果は機械学習分野のトップ国際会議 International Conference on Machine Learning (ICML2016), 機械学習分野のトップ論文誌 Journal of Machine Learning Research において発表するとともに, 多く

の招待講演において報告を行った。

MD 班では、パーシステントホモロジーを用いてタンパク質の立体構造を記述・解析する新たな手法を開発した。これまでタンパク質の立体構造は、Ramachandran's plot や contact map などの手法を用いて記述・解析されてきたが、これらの手法は局所的な構造を元にタンパク質の形状を特徴づけており、アミノ酸が形作る大域的な構造を記述するには不十分であった。これに対し MD 班では、TDA 班で開発された逆解析手法をベースに、タンパク質の立体構造の新たな記述法を提案した。この手法では、パーシステントホモロジー解析で得られた最小生成元の情報を元に、自然言語処理や機械学習手法を併用してタンパク質の特徴量ベクトルを計算する。本提案手法の有効性を見るため、chignolin や HP35(nle-nle)といったタンパク質の折りたたみ過程を分析したところ、従来の方法に比べて容易に準安定状態を同定し、それらの状態の立体構造を特徴づけることに成功した。これにより、位相的データ解析をタンパク質の「記述言語」として使う新たな道筋が開発された。

(2) 顕著な成果

<優れた基礎研究としての成果>

1. パーシステント図の機械学習法の開発

概要:パーシステント図に対する二種類の機械学習法の開発を行った。一つ目はパーシステント図を再生核ヒルベルト空間へ連続に埋め込むカーネル法の理論を構築し、その計算手法の提案やガラス転移点の特徴付けなどの応用研究も行った。また二つ目の線形機械学習法では逆解析も可能となっており、学習結果に対して幾何学的な解釈を与えることができることから、interpretable AI の新たな方法論として注目されている。これらの成果は機械学習の一流誌に掲載され、またこれをもとに多数の企業および大学の材料研究者との共同研究へと発展した (Kusano, Fukumizu, Hiraoka: ICML2016, JMLR2018. Obayashi, Hiraoka, Kimura: JACT2018.)。

2. パーシステント図の確率論的極限定理

概要:パーシステント図に対する確率論的漸近挙動を理論的に解明し、いくつかの極限定理を証明することに成功した。ここで扱った問題は、ユークリッド空間内のランダムな点過程や格子モデルであり、そこから定まるベッチ数の大数の法則と中心極限定理を証明した。またランダム測度理論を用いることで、パーシステント図の極限定理(大数の法則に対応)を証明することにも成功し、ランダムパーシステント図の極限としてあるラドン測度が収束先として得られることを示した。さらにはそのラドン測度の台の特徴付けも行った (Hiraoka, Shirai, Trinh: AAP2018. Hiraoka, Tsunoda: DCG2018)。

3. マルチパラメータ・パーシステントホモロジーの表現論的研究

概要:マルチパラメータ・パーシステント加群の研究に Auslander-Reiten 理論を導入し、従来のパーシステントホモロジー理論を拡張することに成功した。特に、制限付き2パラメータ・パーシステントホモロジーとして可換梯子型クイバーを提案し、梯子の長さに応じた有限性の証明を完成させた。またマルチパラメータ・パーシステント加群に対して、区間分解・区間近似の概念を提案した。この概念は、代数的な取り扱いが困難であったマルチパラメータ・パーシステント加群を実用的なデータ解析技術にするうえで極めて有効な手法になるものと思われる (Escolar and Hiraoka. DCG2016)。

<科学技術イノベーションに大きく寄与する成果>

1. パーシステントホモロジー計算ソフトウェア「HomCloud」の開発

概要:パーシステントホモロジーソフトウェア「HomCloud」を開発した。本ソフトウェアは、世界で初めて逆問題解析、機械学習、スパース解析などのより踏み込んだパーシステントホモロジー

解析手法を組み込んでいる。現在、本ソフトウェアは材料科学を中心に大学および企業の研究現場で広く使われ、TDA を用いた異分野共同研究の基盤を与えている。実際に本 CREST の発展的位置付けとして、国内の材料科学の大型プロジェクト(内閣府 SIP, JST Mirai, NEDO 超超プロジェクト)などで HomCloud は使われており、また最近では材料科学に限らず生命、医療、経済学、環境などの諸分野でも応用されている。また本ソフトウェアを用いて国内材料系企業 6 社との共同研究に発展し、また材料 TDA コンソーシアム「トポロジカルデータ解析コミュニティ」の設立にもつながった(民間企業 19 社、その他研究機関などから 8 機関が参加)。HomCloud Webpage: <https://homcloud.dev>

2. トポロジカルデータ解析の材料科学への応用研究

概要: シリカガラスの原子配置データにパーシステントホモロジーを適用することで、階層的なリング構造が中距離秩序を作っていることを発見した。ここで提案された手法は、シリカガラスに限らず、より広いクラスのガラス、高分子、磁性材料、実験画像解析など幅広い分野へ現在適用されており、多くの材料科学者との共同研究(大学および民間企業)へ展開した。また海外研究グループが我々の取り組みを追従するなど、世界的な評価を受けている(Hiraoka et al. PNAS2016)。

3. トポロジカルデータ解析の横断的応用研究

概要: ソフトウェア「HomCloud」の普及にとめない、材料科学に限定しない多くの異分野共同研究を実施した。医療画像への応用では肝腫瘍の画像解析へパーシステント図の機械学習法を適用し、非常に高い精度で肝細胞癌、転移性腫瘍、肝血管腫の 3 種を識別可能であることが示された。その他にも、生命科学、環境エネルギー問題、企業の特許戦略解析、気象学などへ応用し、論文として成果を発表した(Inatsu et al, SOLA, 2017, Oyama et al, Sci. Rep. 2019, Suzuki et al, Comp. Geosci. 2020, Suzuki et al, Sci. Rep. accepted)。

< 代表的な論文 >

1. Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escobar, K. Matsue, and Y. Nishiura. Hierarchical structures of amorphous solids characterized by persistent homology, Proceedings of the National Academy of Sciences of the United States of America 113, 7035-7040, 2016.

概要: シリカガラスの構造解析にパーシステントホモロジーを適用することで、従来の解析手法では明らかにされていなかった階層的なリング構造を抽出することに成功した。本論文で開発された手法は、極めて複雑な原子配置構造に対して、トポロジーや幾何構造に着目したコンパクトな記述子を与えるものであり、本 CREST 期間中に多くの材料科学研究へ適用されることとなった。また数学的にも、多元環の表現論を用いたパーシステントホモロジーのマルチパラメータ化や、ランダムトポロジーにおける極限定理などへの動機付けを与えることとなった。

2. G. Kusano, K. Fukumizu, and Y. Hiraoka, Persistence weighted Gaussian kernel for topological data analysis, Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, JMLR: W&CP volume 48, 2004-2013, 2016.

概要: パーシステント図のカーネル法によるベクトル化を通じたデータ解析の枠組みを開発した。本論文では Persistence Weighted Gaussian Kernel (PWGK) 法を提案し、既存の手法に比べて本手法によるベクトル化が優れた性質を持つことを、数学的性質および数値実験の双方から示した。また本手法をもとに、シリカのガラス-液相転移点の統計的特徴づけや、粉体の相転移密度付近における特徴的な幾何構造の抽出などに関する成果も得られた。

3. E. Escobar and Y. Hiraoka. Persistence Modules on Commutative Ladders of Finite Type. Discrete and Comput. Geom. 55 (2016), 100-157.

概要: 本論文では多元環の表現としてパーシステントホモロジーの概念を拡張することに成功し、その特別な場合として可換梯子型クイバー ($A_2 \otimes A_n$) を用いたデータ解析手法を提案した。主

結果としては、長さ n が 4 以下の場合には有限表現型, 5 以上の場合には無限表現型になることを明らかにし, また Auslander-Reiten 理論を用いてパーシステント図を一般化した. ここで得られた数学的成果をガラスの構造解析にも応用し, 加圧試験時の微小変形による材料構造の特徴づけにも成功した.

§ 2 研究実施体制

(1) 研究チームの体制について

① TDA班

研究代表者: 平岡 裕章 (京都大学高等研究院 教授)

研究項目

- ・パーシステント逆問題理論の研究
- ・ソフトマターTDA研究
- ・ソフトウェア開発

② 表現論班

主たる共同研究者: 浅芝 秀人 (静岡大学大学院理学領域 教授)

研究項目

- ・有限体上でのbocsの変形
- ・射空間の次元公式
- ・thin直既約加群への直和分解
- ・ソフトウェアの開発

③ 確率論班

主たる共同研究者: 白井 朋之 (九州大学マス・フォア・インダストリ研究所 教授)

研究項目

- ・ランダム複体の研究
- ・点過程およびランダム場のパーシステント図解析
- ・マルコフ過程のパーシステント図解析

④ 統計班

主たる共同研究者: 福水 健次 (統計数理研究所数理・推論研究系 教授)

研究項目

- ・カーネル法によるPDの統計的解析手法の開発
- ・PDブートストラップ法等の統計的信頼度評価法開発
- ・ガウス過程に付随する最大値分布やエクスカージョン過程の研究
- ・アブストラクトチューブの一般化と統計的検定への応用

⑤ MD班

主たる共同研究者: 一宮 尚志 (岐阜大学医学部 准教授)

研究項目

- ・粗視化分子動力学シミュレーションのTDA解析
- ・全原子分子動力学シミュレーションによるタンパク質フォールディングのTDA解析

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

国際的なネットワーク作りとして, アジア太平洋地域の研究者とともに, 2020年夏からオンラインセミナーシリーズ「Asia Pacific Seminar on Applied Topology and Geometry (APA TG)」を立ち上げた. アジア太平洋地域に適した時間帯でオンラインセミナーを実施すること

で、研究活動の活性化、若手研究者の育成、新たな人材の発掘などを行なっている。世界各地から240名が参加登録をしている(2021年8月時点)。

APATG Webpage: <https://sites.google.com/kyoto-u.ac.jp/apatg>

産業界との連携としては民間企業向けコンソーシアム「トポロジカルデータ解析コミュニティ」を2021年2月に立ち上げた。このコンソーシアムでは、本CRESTで開発した材料TDAに関する手法を広く産業界へ還元することを目指しており、民間企業19社、その他研究機関などから8機関が参加している(2021年8月時点)。活動内容としては定期的に解析手法やソフトウェアHomCloudのチュートリアルを実施し、また企業側からの個別案件についてもコンサルタントを行なっている。今後はそこで得られたテーマなどをもとに、共同研究に展開させることも目指している。

Webpage: <https://www.wpi-aimr.tohoku.ac.jp/TDA/>

また研究実施中間報告書にも記載したが、本CRESTチームでは開発した研究成果を材料科学を中心とした異分野共同研究へ積極的に展開している。まずTDA班では、物質・材料研究機構、高エネルギー加速器研究機構、さらには民間企業との共同研究や、SIP革新的構造材料(内閣府)、情報統合型物質・材料開発イニシアティブMI²I事業(JST)、超先端材料超高速開発基盤技術プロジェクト(NEDO)などの大型材料科学プロジェクトとの連携による実用化展開を並行して進めた。これらの大型プロジェクトおよび企業共同研究の活動により、我々が開発したソフトウェアHomCloudは、現在多くの材料科学者から利用されている。また最近では本CREST成果をもとに、生命科学(遺伝子データ解析)、医療画像解析、経済時系列モデルといった分野との共同研究も実施しており、数学の強みである横方向の展開も行われている。これらに加えて、確率論班では九州大学I2CNERの辻健氏と山口大学工学部 Fei Jiang氏との共同研究で、天然の岩石の不均質な細孔の幾何学的形状を、パーシステント図を用いて考察している。また統計班ではCREST研究チームの成果展開による自動車メーカーとの共同研究を実施している。