公開

研 究 終 了 報 告 書

「パーシステントホモロジーによる位相高次構造抽出手法開発」

研究期間： 2019 年 10 月～2023 年 3 月

研 究 者： 大林一平

1. 研究のねらい

本研究計画の大きな目標は，パーシステントホモロジー(PH)によるデータ解析プラットフォームの発展である．数学のトポロジーの概念を利用したデータ解析(これを TDA と呼ぶ)が 21 世紀に理論からソフトウェア，応用まで急速に発展している．PH は TDA の主要ツールで，材料科学や生命科学のデータ解析などに利用が広がりつつある．PH はデータの形の情報をパーシステンス図(PD)という形で縮約し，解析する．プロジェクト開始の時期は PH の基礎理論が確立し様々な応用が広がり始めた時期であった．応用が進むにつれて，単に PD が計算できるだけでは不十分で，様々な周辺的な道具が必要がことがわかりはじめた．応用側から必要とされる道具を数学/数理方面からのアプローチによって開発することで，PH によるデータ解析の領域をさらに豊かにすることができるだろうというのがこのプロジェクトの目論見である．

具体的には，以下の 3 つの小目標に分解し，これらの結果を融合することで PH によるデータ解析の発展を実現することが本研究プロジェクトでの作戦である．

- 数学的理論，数理的手法の開発
- データ解析ソフトウェアの開発
- 以上の手法，ソフトウェアを利用した応用(データ解析)

理論や手法の開発のためには，ホモロジー論の様々な道具の活用，数理最適化の活用，機械学習と PH の組み合わせ手法の開発，などを目標達成のための具体的手段として計画した．ソフトウェアの開発としては，既に開発を進めていた PH に基づくデータ解析ソフトウェア HomCloud を改良していくこと計画した．応用については，材料科学データへの応用を中心に計画していた．特に分子動力学シミュレーション等で得られた原子配置データや各種顕微鏡で観測した材料の実データなどの解析を実施する予定であった．本研究計画の重要なこととして，理論/手法→ソフトウェア→応用と一方向的に進めるのでなく，応用→ソフトウェア→理論/手法というフィードバックを重視することで，より有用なソフトウェア，より有用な理論，を開発することを目指す点である．この双方向的な進め方は，理論/ソフトウェア/応用の全てに経験，強みがある大林ならではのやりかたである．

2. 研究成果

(1)概要

まず, 理論/手法研究に関しては以下の成果を上げることができた.

- stable volumes
- PH の体の選択問題
- NMF と PH の組み合わせ手法の開発(と手法の焼結鉱の 3 次元 X 線 CT 画像への適用)

応用研究に関しては次のような成果がある.

- 材料科学への応用
  - 金属ガラスの中距離秩序の PH による特徴付け
  - Si ガラスの熱伝導率と形の関係を PH で評価
  - PH によるアモルファスの機械学習ポテンシャルの構築
- 地質学への応用
  - 岩石の透水率の PH による予測

上で挙げたような応用研究は基本的にすべて応用分野における研究者との連携によるものである. さらにこういった連携研究を強化するため, TDA-MI workshop と題した研究集会を 2 回実施した. その他トポロジカルデータ解析コミュニティの立ち上げといった活動も行い, PH の普及活動に邁進した. ソフトウェア開発に関しては HomCloud のインストールの容易化や利用環境の拡充, 新機能の追加や高速化, ドキュメントの拡充などを行い, 継続的開発とリリースをすることができた.

全体としては, 理論, ソフトウェア開発, 応用をバランス良く進めることができたように思われる. 応用から理論へのフィードバックという目論見も, stable volume や NMF+PH でうまく機能したと言える.

またソフトウェアの改良やドキュメントの充実などの効果として, 大林に相談しなくても自発的に HomCloud でデータ解析できる人が増えているのではないか, という実感がある. 実際大林が関係してない研究でも HomCloud の利用が増えており, 知る限りでも 20 件以上の事例(論文, プレプリントなど)が確認されている. 応用研究で挙げた Si ガラスの熱伝導率の研究や機械学習ポテンシャルの件についても共同研究者からの相談が来た時点でかなり HomCloud を使いこなしており, 大林の作業がボトルネックになることなく研究を進めることができた. HomCloud は研究活動の「てこ」として有用なソフトウェアへと進化していると言える.

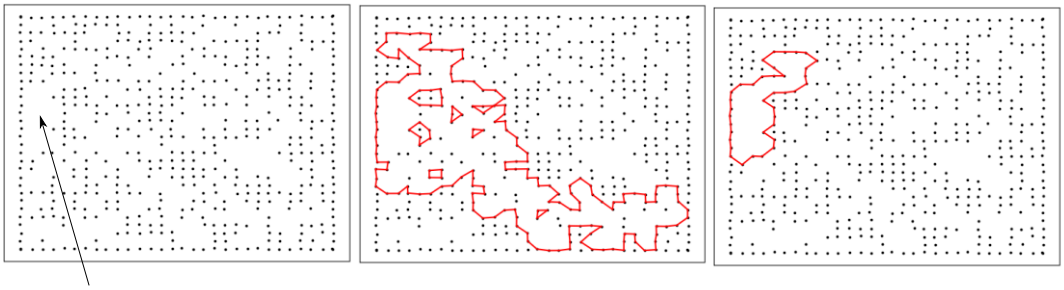
(2)詳細

以下, 概要で紹介したさきがけ研究期間における研究成果を項目別に記載する.

Stable volumes (代表的な論文 1)

PH を用いて得られる PD は平面上の散布図で, その各点はデータ上の連結成分やリング構造, 空隙構造などに対応している. そこで PD の点が元データのどの構造に対応しているかを特定することができれば PH によるデータ解析において非常に有用な道具

となる．これは PH の逆解析と呼ばれるもので，ホモロジー代数上の数理最適化によって様々な手法が提案されている．ただ，既存の手法には (1) ノイズへの弱さ (2) 最小の構成要素の抽出に確率的に失敗する，という問題があり，結果として利用者の直感に合わない結果を返す場合があった．これは実践的にも結晶的構造を持つ材料データの解析で生じる問題である(下図).



大林は既存手法の一つである optimal volume を改造した stable volume とその変種である stable sub-volume をこの問題の解決策として提案した．特殊な場合に 2 種類の定式化を準備し，この 2 つが同値であることを証明した．またこの定式化がノイズに対して変動しない部分に対応していることを証明し，ノイズに対する性質を数学的に保証することに成功した．そして一般的な場合もこの定式化が利用可能であることを示した．Stable volume を計算するプログラムの実装し，人工データ（上図，中央が既存手法の結果で右が stable volume の結果）や材料科学のデータなどに適用することでその機能を確認した．この手法は PH によるデータ解析に今後便利に活用できることが期待される．

この手法は逆解析の結果が時々予想外の結果となる，という利用者からの要望を解決するために開発された機能である．この意味で stable volume は応用からのフィードバックがうまく手法開発に繋がったと言える．また，この機能は理論ができる前に実装がなされたという経緯がある．つまり理屈はわからないが直感的にこういうプログラムを書けば上手く行きそうという確信のもと実装され，実際にうまく行くプログラムが得られたのである．この意味で stable volume は HomCloud というプラットフォームがあってこそ実現したと言える．「研究のねらい」で述べた応用→ソフトウェア→理論というルートがうまく機能した成果であると言える．

PH の体の選択問題（論文は Discrete Comput. Geom. にアクセプト済み）

この研究は PH において係数体の Z/pZ の素数 p を変えたり， R に変えたり場合にどのような現象が起きるか，係数体の変化に PD が影響されない条件は何か，その条件を判定する効率的なアルゴリズムはないか，という疑問に答えることを目標としている．成果としては，(1) 係数体の変化に PD が影響される十分に良い必要条件や十分条件を Z 係数相対ホモロジーの捩れ部分群の言葉で記述した (2) この条件を効率的に判定するアルゴリズムを提案し，HomCloud に実装した (3) ランダム生成した単体複体を使って「係数体の変化が PD に影響する頻度」を数値的に見積もった，などが挙げられる．例えば三次元データであればこの頻度は十分低く，この問題をあまり気にする必要はないというこ

とが示唆される結果が得られた. 一方高次元データでは高頻度でこの依存性が生じるため, もっとシリアスにこの問題を考えるべき，という示唆も得られた.

この体の選択問題は普段はあまり気にされていないが，数学に強い人には比較的に気になっていた問題であり，これを上手く解決した成果である.

NMF と PH によるデータ解析手法の開発とこの手法による焼結鉱の 3 次元 X 線 CT 画像の解析（代表的な論文 2）

この研究は，焼結鉱の 3 次元 X 線 CT 画像の解析の PH による探索的データ解析のために新しいツールを開発したという物である．数学的な理論から，１つのデータからはデータの次元と同じ数の PD が得られる．この複数の PD の情報を統合し分析するための手法として，(1) 非負行列分解の利用 (2) PD から得られる特徴量ベクトルの連結，という２つのアイデアを考案し，焼結鉱の 3 次元 X 線 CT 画像を解析した．結果としてこのデータには３種類の特徴的構造があることを発見した.

１つのデータから得られる複数の PD の情報の統合というのは PH によるデータ解析の課題の一つで，この問題に一つの回答を与えた結果であると言える．特徴量ベクトルの連結というやり方はかなり安直な手法であるが，安直な手法が良い解析結果を生み出せるという点で面白い成果であると言える.

HomCloud の開発

HomCloud の開発については，まず (1) インストールの簡易化 (2) 対応環境の増加 (3) ドキュメントの充実，という利用者からの利便性の向上という意味で大きな進展があった．(1) については，Python の標準的なパッケージ管理システム pypi への対応や Windows 用バイナリパッケージの提供などが挙げられる. これらによってインストールがかなり手軽になった．(2) については Apple Silicon Mac への対応，Anaconda や Google Colab といったデータ解析専門プラットフォームへの対応, などが成果として挙げられる. 特に Google Colab への対応によってデータ分析に馴染みがない者でも HomCloud による PH 解析を体験できるようになったという意味で PH の普及に一役買う進展である．(3)については，インストールガイドやチュートリアルの拡充，またユースケース（利用実績リスト）の充実などが挙げられる．　機能の追加やインターフェースの改良，高速化なども上の改良と並行して行った．この時期に行った顕著な改良として，2,3,4 次元の立方体複体（画像データ）の大幅な高速化などが挙げられる．　こういった改良と並行して内部構造の改善も随時進めた．内部構造の改善は今後の機能拡張や高速化の基盤となるもので，こういった改良を随時行ったのはプロジェクト終了後の発展にも重要である.
こういったソフトウェアの改良やドキュメントの充実の結果として，HomCloud を利用した研究も増えつつある．大林が関係していない研究事例でも HomCloud を利用した論文やプレプリントが 20 本以上ある．材料科学(鉄，ポリマー，高分子材料，粉体ガスなど), 生命科学(計算機援用診断や発生生物学)，地質学，量子色力学，計算流体力学など多様な

分野で HomCloud の利用が広がりつつある．詳しくは https://homcloud.dev/use_cases.html を参照して欲しい．

材料科学や地質学への応用（代表的な論文 3 など）

材料科学への応用としては，(1) 金属ガラスの中距離秩序の PH による特徴づけ (2) Si ガラスの熱伝導率と形の関係を PH で評価 (3) PH をアモルファスの機械学習ポテンシャルの構築，などが挙げられる．また上で述べた三次元 X 線 CT 画像の解析は手法の提案と材料科学への応用の 2 つの側面を持つ研究である．地質学への応用としては岩石の構造と透水率を PH 経由で関連付ける研究を行った．　これらの研究は基本的な枠組みは似ていてデータから PD を計算し，機械学習の手法と PH の逆解析を組み合わせることで材料の物性や秩序構造と関係のある局所的な構造を抽出することが基本となっている．ただ，ここから先はそれぞれの領域のドメイン特有の知識を用いて解析結果の意味づけをする必要があっ．金属ガラスの場合は構造因子という指標が中距離秩序の特徴づけという目標に有効であったし，Si ガラスの熱伝導率の研究については local vibration mode (局所振動モード)を用いた解析を併用することが物性との関連を明らかにするのに重要であった．こういった検証は材料科学の専門家が必要で，数学外とのコラボレーションがうまくいった成果である．

アウトリーチ活動やワークショップの開催など

材料科学と PH の連携研究をさらに推進するため，2020 年 11 月と 2022 年 12 月に TDA MI workshop という名前で研究会を開催した．2020 年度はオンライン，2022 年度はハイブリッド形式で開催した．2020 年度は 2 日で 13 件，2022 年度は 1 日で 7 件の発表が行われ，盛況であった．また，東北大 AIMR の赤木氏，京大の平岡氏とともにトポロジカルデータ解析コミュニティという位相的データ解析の材料科学への応用に関心がある人々のためのコミュニティの立ち上げを行った．　この他にも，HomCloud を用いたデータ解析に関する企業とのコンサルティングを実施したり，HomCloud の利用法に関する国際チュートリアルを実施するなどアウトリーチ活動や企業との連携活動なども実施し，本プロジェクトの成果を外部に広めることに努力した．

3. 今後の展開

研究の今後の展開について

機械学習+PH や数理最適化+PH という方面では，topological loss function という PH の情報を損失関数にする研究がなされており，有望な応用がありそうである．HomCloud でも topological loss function を微分する機能(これはこの損失関数を最小化するのに必要となる機能である)を実装しうまく動くことを確認したので，このアイデアを利用した研究を

進めたい.

研究成果の将来的な社会実装に向けて

本プロジェクトの成果は PH によるデータ解析フレームワークの発展であるので，これだけでは社会実装にまで繋がるわけではない．こういったデータ解析の仕組みを研究者やデータサイエンティストが活用することで社会へと貢献できると考えられる．本プロジェクトでは理論や手法の開発だけでなく HomCloud の開発やドキュメンテーションの充実などに努力してきたが，そういった活動がより社会実装に繋がるものと期待される．

大林による PH の応用先は材料科学が中心であるので，今の研究からどの程度の期間あれば企業からの材料製品になるのか検討してみる．これまでの応用研究で材料系の企業の研究者に興味を持ってもらう，という所には既に来ている．2022 年度に企業の研究者と研究開発に関するコンサルティングをした経験では，そういった人々が PH を使いこなせるには半年〜1 年程度必要そうである．PH の材料科学への応用は当たり外れが大きく，うまくはまるターゲットを見付けるには 1，2 年は必要だろう．材料系では研究から製造まで行くのに 5 年 10 年というスパンが必要なようである．つまり現状から社会実装に行くまでに少なくとも 10 年は必要そうである．一方で材料科学の世界では計測技術やシミュレーション技術の発展によりデータ量が急速に増えているという背景があり，PH がそういったデータの解析に喰い込める可能性は十分ある．大学の研究者としてはじっくりと研究するのが良さそうである．

4. 自己評価

研究目標の達成については，かなり良く進展したと思われる．客観的な数字としても発表論文 6 本，アクセプト済み論文 2 本，査読中論文 1 本，発表 30 件，書籍 1 件，総説等 3 件などがあり，HomCloud も 3.0.0 のリリースから 3.6.0 までバージョンアップすることができた．本格的な PH の教科書の執筆も進行中で，2023 年中には出版する予定である．応用に有用な理論を開発したい，という目標に関しては stable volume や NMF の研究などがうまくいったと考えられる．

研究実施体制等については，HomCloud の技術支援のためにスタッフを雇用したことはソフトウェアの構築やユースケースのリスト整備，ドキュメンテーションの拡充等に有効であった．また学生の RA に HomCloud のプログラミングを手伝ってもらったのも比較的うまくいったと思われる．学生の RA は 1 年だけで，もっと活用できれば良かったとは思われるが，これは人の巡り合わせの問題で運が大きく絡んでくる問題であるし，岡山大への異動もあり難しかった．

研究成果の波及効果については HomCloud の開発や周辺的な環境整備が今後のデータサイエンスの発展に一定の効果があるのではないか，と期待している．PH は「適用範囲は狭いがうまく嵌ると非常に良い結果を出す」道具であるので，PH に気軽に手を出せる環境の整備が重要であるはずである．

5. 主な研究成果リスト

(1)代表的な論文(原著論文)発表

研究期間累積件数:6件 (さらに2件が期間中にアクセプトされ, 1件が査読中)

1. Ippei Obayashi. Stable Volumes for Persistent Homology. Submitted to Applied and computational topology. Journal of Applied and Computational Topology, online first article, (2023) (研究期間中にアクセプト, 終了後に掲載)
本論文は PH の逆解析の新しい手法, stable volume とその変種 stable sub-volume を提案した. 既存の手法の弱点である (1) ノイズに弱い (2) 最小の構造の抽出に時々失敗するという問題(これは特に結晶的構造の解析で問題となる)を解決するための手法である. stable volume の定式化と計算アルゴリズム, 数学的性質の証明, 人工データおよび実データへの適用例の紹介, などを行い手法の数理的性質の良さ, および実際的な実用性を示した.
2. Ippei Obayashi, and Masao Kimura. Persistent homology analysis with nonnegative matrix factorization for 3D voxel data of iron ore sinters. JSIAM Letters **14**, 151–154 (2022) https://doi.org/10.14495/jsiaml.14.151
本論文は PH と非負行列分解(NMF)を組み合わせた手法を提案し, 焼結鉱の 3 次元 X 線 CT 画像へと適用した結果に関するものである. NMF と組み合わせる際の PD のベクトル化で次元の異なる PD のベクトルを並べて連結するというちょっとした工夫を組み合わせることで特徴的な共存構造を発見することを可能とした. この手法を焼結鉱の 3 次元に適用することで 3 つの特徴的共存構造を特定するとに成功した.
3. Emi Minamitani, Takuma Shiga, Makoto Kashiwagi, and Ippei Obayashi. Topological descriptor of thermal conductivity in amorphous Si. **156**, 244502 (2022) https://doi.org/10.1063/5.0093441
本論文はアモルファスシリコンの熱伝導率とその原子配置の関係を調べるために PH を利用したものである. アモルファスシリコンの原子配置を分子動力学シミュレーションで計算し, 熱伝導率をその配置から計算した. さらに原子配置から計算した PD を記述子, 熱伝導率を目的変数とする回帰を実施することで熱伝導率に効く局所的な構造を特定し, その構造をさらに解析することで物理的意味付けを得ることに成功した.

(2)特許出願

研究期間全出願件数:0 件(特許公開前のものも含む)

1	発　明　者	
	発明の名称	

	出　願　人	
	出　願　日	
	出願番号	
	概　　要	
2	発　明　者	
	発明の名称	
	出　願　人	
	出　願　日	
	出願番号	
	概　　要	

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- ソフトウェア HomCloud のリリース

本プロジェクトの大きな目標である HomCloud の開発成果のリリースは，期間中十分な頻度で実施することができた (3.0.0, 3.1.0, 3.2.0, 3.3.0, 3.4.0, 3.5.0, 3.5.1, 3.5.2, 3.6.0). 期間後に 4.0.0 のリリースが実施され，今後もバージョンアップが続く予定である.

- 教科書の執筆

PH に関する教科書の執筆を行った. マテリアルズインフォマティクスに関する教科書の 1 パートの執筆を東北大 AIMR 赤木氏とともに行い，2022 年度に出版された. こちらは主に材料科学者に向けた内容となっている. また，PH の専門家4人で PH の理論と応用に関する幅広い内容を扱った教科書を現在執筆中で 2023 年中の出版を予定している. 後者は理論家から応用に興味のある人に幅広く PH をアピールできる内容となる予定である.

- PH に関する講演

研究成果に関する講演はもちろん，PH の概要，既存の応用の紹介，HomCloud の宣伝といった内容の講演も期間中に何度も行った. これは PH の利用の拡大や HomCloud の普及に大きなプラスになると期待される. また今後の材料科学などへの応用の共同研究にも繋がると期待される.

- 第 11 回 桜舞賞 研究奨励賞(Development of topological data analysis software and its industrial applications)

本賞は理化学研究所の研究者で，活発な研究活動を行い、優れた研究成果及び顕著な貢献のあった研究者に授与する賞である. HomCloud の開発およびその材料科学への応用が評価されてこの賞を受けるに至った.

- TDA MI Workshop の開催

本ワークショップは位相的データ解析とマテリアルズインフォマティクスの互いの研究成果や知見を交換し，相互のコラボレーションを促進するために実施された. 2020 年(オン

ライン)と 2022 年(ハイブリッド)の 2 回開催し, ともに活発な発表, 議論が行われた.