

位相的データ分析法による材料構造形成過程の分析

Topological Data Analysis of Domain Pattern Formation in Materials

本 武 陽 一*・水 牧 仁 一 朗**・工 藤 和 恵***・福 水 健 次*
Yoh-ichi MOTOTAKE, Masaichiro MIZUMAKI, Kazue KUDO and Kenji FUKUMIZU

(Received 15 January 2021, Accepted 26 February 2021)

Key Words: Topological Data Analysis, Persistence Homology, Pattern Formation, Time Dependent Ginzburg Landau Equation, Reduced Model

1. はじめに

物理学の歴史において、ハミルトン力学のような座標系の取り方によらない運動理論は、量子力学といった様々な現象を統一的に扱うための理論的な基盤となった。また、結晶構造のような周期構造に対するフーリエ基底空間や、熱力学現象のようなランダム現象における圧力や温度といった統計量空間は、それらの重要な情報を保持しつつ様々な現象を統一的に理論化する上での基礎を与えた。一方で、構造材料における転移パターンの形成と発達のような、非周期的な秩序構造をもつ、非線形で非平衡なパターン形成現象を一般に取り扱える特徴量空間は未だに発見されていない。また、特徴量空間の取り方によって分析結果が大きく変わるとい現象は、物理学に限らず、画像やネットの購買履歴データといった一般のデータ分析で頻繁に生じる現象でもある。したがって、このような特徴量空間の設定は、幅広い分野のデータ分析においてとても重要な課題である。

位相的データ分析 (Topological Data Analysis: TDA) の提唱者の1人である Carlsson 博士は、位相的データ分析を行う動機の一つとして、座標系の取り方にできるだけ依存しない特徴量空間の設定を挙げている¹⁾。トポロジーとは、連続な変形によって移りあえる図形を全て同一視することで、対象の「かたち」をカテゴライズするという大胆な粗視化を行った特徴量のことである。Carlsson 博士はこのようなトポロジーの性質から、座標系の違いのような入力空間の差異に影響されない特徴量の設定が可能になると考えたのである。

一方で、空間的なスケールやオーダーといった情報は、物理現象を理解する上で欠くことのできない情報である。このような要求に応える特徴量として、位相的データ分析

にはパーシステントホモロジーという特徴量が用意されている。パーシステントホモロジーは後で詳述するように、トポロジカルな情報にその構造のスケール等と関連した情報を付加した特徴量となっている。言い換えれば、パーシステントホモロジーは、スケール情報を適度に含みつつ、大胆な粗視化を行った特徴量とみることができる。このようなパーシステントホモロジーを用いることによって、例えばシリカガラスのアモルファス状態と、液体状態の違いを特徴付けることができることが報告されている²⁾。これは、非周期的な秩序構造を持つアモルファスの準長距離秩序をパーシステントホモロジーによって特徴付けられることを意味している。そのため、多くの非周期的な秩序構造をもつ材料構造の特徴づけにおいても、パーシステントホモロジーによる分析が有用であることが示唆される³⁾。

我々は、このパーシステントホモロジーという特徴量によって構成される空間が、非線形で非平衡なパターン形成過程を統一的に記述するための基盤となるのではないかという思想のもとに、パーシステントホモロジーを用いて物理・工学的な発見・発明を試みる研究を行っている。本稿では、最近我々が試みている材料構造やその形成過程のパーシステントホモロジーを用いた分析についての事例紹介を行うとともに、著者が位相的データ分析を始める上で得た知見を紹介できればと考えている。

本稿のパーシステントホモロジーの説明は、既に位相的データ分析を始めたが、そこで算出される特徴量の具体的な算出原理がわからず、納得感を得られないような研究者を想定した。また説明は、概論的な説明と難解な数学的な公理・証明型の説明の中間に位置するような、応用を行う上で手法を納得・安心して使うために最低限必要と思われる簡易的な説明のみを行った。したがって、より精緻な理解のためには、

* 統計数理研究所 (〒190-8562 東京都立川市緑町 10-3)

The Institute of Statistical Mathematics (10-3 Midori-cho Tachikawa, Tokyo 190-8562, Japan)

** 公益財団法人高輝度光科学研究センター (〒679-5148 兵庫県佐用郡佐用町光都 1-1-1)

Japan Synchrotron Radiation Research Institute (1-1-1, Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5198 Japan)

*** お茶の水女子大学基幹研究院 (〒112-8610 東京都文京区大塚 2-1-1)

Department of Computer Science, Ochanomizu University (2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan)

引用した参考文献などを適宜参照されたい。また、そのような解説目的を設定した結果、位相的データ分析の研究事例紹介より位相的データ分析の説明に頁数が割かれている。この点に注意して本稿を読み進めてもらいたい。

2. 位相的データ分析

位相的データ分析とは、データの位相幾何学的な構造に着目したデータ分析法のことである。位相的データ分析は、結晶の原子座標データや、スペクトル強度、原子・分子の密度分布データ、その他各種計測画像データといった、構造を持つデータを対象として、穴やネットワーク構造といったトポロジカルな特徴を抽出する。一般にトポロジカルな情報は、空間のスケールやオーダーといった物理現象を理解する上で欠くことのできない距離情報を持たない。本稿で紹介するパーシステントホモロジーでは空間のスケール情報を付加することにより物理的に有用な情報を適切に取り出すことが可能である。そこでまず、数学的な正確性には目を瞑ってもらった上で、応用者である著者が手法を使用する上で必要と考える範囲内の、具体例に即したパーシステントホモロジー群についての説明を試みる。その上で、なぜパーシステントホモロジー群を用いた解析が物質構造の分析に有用であるのか理由を述べる。解説は参考文献^{4),5)}などに基づく。

2.1 入力データと構造の代数表現

磁性体のドメイン構造や、結晶・構造材料に見られる穴構造は、その物性的・力学的特性を決定する主要な要因となる^{6),7)}。その特徴は、まさに穴の数という位相幾何的な特徴量によって抽出される可能性があると考えられる。このような穴を数学的に厳密に定義したものをホモロジー群と呼ぶ。位相的データ分析では、足し引きが可能となるように図形を代数的に表現することで、このホモロジー群をなんと線形代数の演算によって算出する。

ホモロジー群の算出のために、まずはパターン構造を代数的な操作が可能な表現に変換する必要がある。そこで重要となるのが脈体定理である。脈体定理とは、円や四角形

のように凹みのない閉じた図形(凸閉集合)の有限個の集まりで被覆された構造があった場合、その構造はトポロジカルな特徴を保存(ホモトピー同値)しながら、代数操作が可能な単体複体による表現に変換できるというものである。ちなみに、ピクセル画像データは方体複体を用いても同様な表現ができる。方体/単体複体とは、四角形あるいは三角形の組み合わせによって図形を表現したもののことであり、これを用いてホモロジー群を算出する計算を行う。以下ではこの方体/単体複体をどのように構成するかを具体例を用いて説明する。

密度分布や測定画像といった、ピクセル画像として得られる実験・数値計算データがもつドメイン構造(Fig. 1(a))や、結晶構造における原子座標や、粒子シミュレーション、高分子構造といった点群として得られるデータの分布(Fig. 1(b))を、凸閉集合の和集合として表現することを考える。まず、ピクセルデータは、一つのピクセルを正方形とした凸閉集合の和集合として考えることができる(Fig. 1(b)上段)。また、点群データでは、例えば点を中心とする球の空間充填構造や格子空間中の欠陥構造などが重要な特徴となることがあるため、点を中心とした半径 r の円や球による凸閉集合の和集合を考えることが有用と考えられる(Fig. 1(b)下段)。

ターゲットのパターン構造を閉凸集合の和集合として表現できたら、次はそれを方体/単体複体に対応づける(Fig. 1(c))。方体/単体複体は四角形/三角形を組み合わせた構造であると述べたが、実はそれには一般に、 d 次元の方体/単体複体が定義される。つまり、0次元の方体/単体は点/点を、1次元は辺/辺を、2次元は四角形/三角形を、3次元は立方体/四面体を現す。ちなみに、組み合わせる際には、異なる方体/単体複体の頂点同士が一致するように“きちんと”張り合わせることも方体/単体複体となるための条件となる。方体/単体複体の構成法としては、例えばピクセルデータの場合、白がドメインだと考えると、白いピクセルの中心には0次元の方体(点)があり、白いピクセルが隣り合えば1次元の方体(線)となり、さらに、隣り合う 2×2 のピクセルが白い場合は、2次元の方体(四角形)を構成

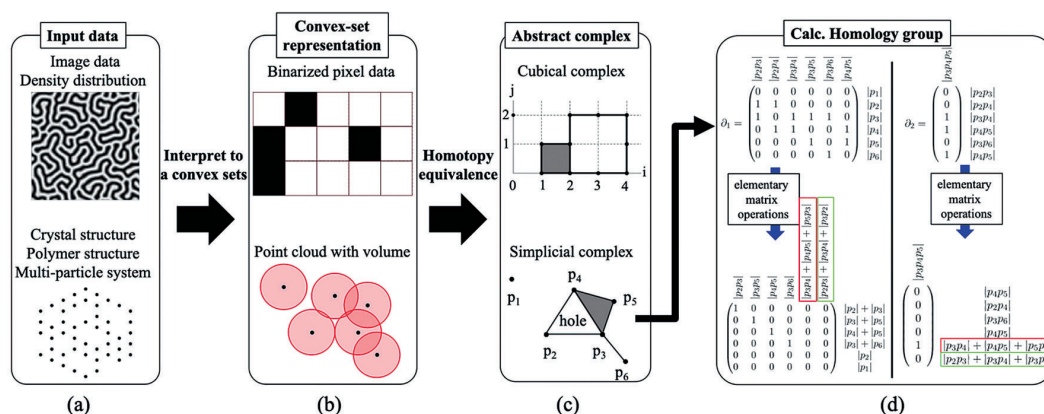


Fig. 1 Procedure of calculation of homology group (hole).

するとする。これによって例えば、**Fig. 1 (b)** の上段にあるようなパターンは、**Fig. 1 (c)** の上段のような方体複体として表現できる。ちなみに、**Fig. 1 (c)** 上段におけるグレーの領域は、ここに2次元方体(四角形の面)があることを意味している。

ここで、**Fig. 1 (c)** 上段にあるようなグリッド空間のある座標点 (i_1, j_1) にある0次元方体(点)を、 $[i_1] \times [j_1]$ と表現することとする。同様に、 (i_1, j_1) と (i_1, j_2) を結んでできる1次元方体(直線)を $[i_1] \times [j_1, j_2]$ 、 (i_1, j_1) と (i_2, j_1) を結んでできる1次元方体と、 (i_1, j_1) と (i_1, j_2) を結んでできる1次元方体によって張られる四角形(2次元方体)を $[i_1, i_2] \times [j_1, j_2]$ と表現することとする。これらの表現を用いると、**Fig. 1 (c)** 上段の方体複体は、次のような方体の集合として表現される。

$$\begin{aligned}
 K_{\text{cubical}} = & \{[0] \times [2], [1] \times [0], [1] \times [1], [2] \times [0], [2] \times [1], \\
 & [2] \times [2], [3] \times [0], [3] \times [2], [4] \times [0], [4] \times [1], [4] \times [2], \\
 & [1, 2] \times [0], [2, 3] \times [0], [3, 4] \times [0], [1, 2] \times [1], [2, 3] \times [2], \\
 & [3, 4] \times [2], [1] \times [0, 1], [2] \times [0, 1], [4] \times [0, 1], \\
 & [2] \times [1, 2], [4] \times [1, 2], [1, 2] \times [0, 1]\}
 \end{aligned}$$

立方体のような3次元構造や、より高次元(k 次元)の構造に対応する方体を考える場合は、 k 方体: $[I_1, J_1] \times [I_2, J_2] \cdots [I_k, J_k]$ を考えれば良い。

同様に、**Fig. 1 (b)** 下段のような点群データは、**Fig. 1 (c)** の下段の単体複体とみなせる。そしてその単体複体は、0次元単体(点)を $|p_i|$ 、1次元単体(線)を $|p_i p_j|$ 、2次元単体(三角形の面)を $|p_i p_j p_k|$ とすることで、以下のような単体の集合として表現される。

$$\begin{aligned}
 K_{\text{simplicial}} = & \{|p_1|, |p_2|, |p_2|, |p_4|, |p_5|, |p_6|, \\
 & |p_2 p_3|, |p_3 p_4|, |p_3 p_5|, |p_3 p_6|, |p_4 p_2|, |p_4 p_5|, |p_3 p_4 p_5|\}
 \end{aligned}$$

このような単体複体をもつ「かたち」を単体の足し算・引き算として表現することによって、数学的に定義された穴であるところのホモロジー群が算出される。ここで注意されたいのは、ホモロジー群では各次元の穴が考えられるということである。例えば、0次元の穴は切れ目なく繋がった単体複体の塊である連結成分であり、**Fig. 1 (c)** の2例には2つの穴があることになる。また、1次元の穴は1単体が繋がってできるリングであり、**Fig. 1 (c)** における対応物は次節で解説する。3次元以上の空間におけるデータでは、さらに高次元の穴も考えることができ、例えば2次元の穴は2単体の組み合わせによって構成される空洞と定義される。ホモロジー群も、それぞれの穴の次元に合わせて0次ホモロジー H_0 、1次ホモロジー H_1 、2次ホモロジー H_2 のようにそれぞれ定義される。次節ではリング状の穴に対応する H_1 の算出法の詳細を説明する。

[補遺] 最後に注意されたいのは、分析ターゲットのパターンを閉凸集合で覆い尽くせない場合、方体/単体複体と元の図形のホモロジー群が一致する保証はないということである。一方で、計算量的困難から、この一致を一部緩和して単体複体を構成するリップス複体⁵⁾ という手法もある。この手法は広く使われているが、特に微細なスケールの穴

の構造に着目する場合、閉凸集合とならない影響が生じるため、分析結果の解釈には注意が必要となる。

2.2 ホモロジー群の代数演算による算出

本節では、前節で得られたターゲットパターンの方体/単体複体による表現を元にして、1次のホモロジー群 H_1 を代数演算によって算出する方法を説明する。

方体/単体複体で構成された構造(**Fig. 1 (c)**)をみると、1次元方体複体の部分集合 $\{[2, 3] \times [0], [3, 4] \times [0], [2, 3] \times [2], [3, 4] \times [2], [2] \times [0, 1], [4] \times [0, 1], [2] \times [1, 2], [4] \times [1, 2]\}$ や、1次元単体複体の部分集合 $\{|p_4 p_3|, |p_3 p_2|, |p_2 p_4|\}$ が1次元の穴(リング)となっていることがわかる。では、同様に存在する1次元方体複体の部分集合 $\{[1, 2] \times [0], [1, 2] \times [1], [1] \times [0, 1], [2] \times [0, 1]\}$ や1次元単体複体の部分集合 $\{|p_3 p_4|, |p_4 p_5|, |p_5 p_3|\}$ で囲まれた領域は穴ではないのだろうか。閉凸集合で構成された元の図形(**Fig. 1 (b)**)に戻るとこれが穴でないことは容易にわかる。対応する領域は正方形や円によって占められていて、穴は空いていないのである。**Fig. 1 (c)** ではこのことはグレーで着色することによって表現されており、この穴の候補となる場所は正方形 $[1, 2] \times [0, 1]$ や三角形 $|p_3 p_4 p_5|$ という面で覆われているため、穴が塞がっているのである。このように、穴とは方体/単体をつないで一周繋がるような「サイクル」であり、かつより高次の方体/単体で塞がれた領域の「境界」とならないものと言える。つまり、「サイクル」と「境界」を算出できれば穴を抽出することができるのである。

ここで、この「サイクル」と「境界」を、方体/単体複体を用いてどのように表現するかについて考える。まず「サイクル」について検討する。先ほど考察したように、サイクルとは一周回って元に戻るような構造のことである。それはすなわち、該当する図形の単体複体が切れ目なく繋がっていて、「境界」を持たないことと等価である。つまり、「サイクル」を算出するには、実は「境界」を算出すれば良いのである。以降、紙面の都合で、単体複体のケースのみを用いて、「境界」の算出方法を説明する。

単体複体における境界とは、例えば1次元単体 $|p_2 p_3|$ であれば、辺の両端点(0次元単体) $|p_2|, |p_3|$ がその境界となる。2次元単体 $|p_3 p_4 p_5|$ であれば、それは面を囲う3辺(1次元単体) $|p_3 p_4|, |p_4 p_5|, |p_3 p_5|$ となる。そして、境界が切れることなく繋がるとは、1次元単体で構成される $\{|p_2 p_3|, |p_3 p_4|, |p_4 p_3|\}$ のような集合で、その各単体の端点が過不足なく繋がることを意味する。つまり、各辺の端点を基底とした足し算を考えると、サイクルであるためには、 $(|p_2| + |p_3|) + (|p_3| + |p_4|) + (|p_4| + |p_3|) = 2|p_2| + 2|p_3| + 2|p_4|$ となるように、図形の端点が2の倍数個存在することが要請される。これより、過不足なく端点が繋がってサイクルを構成するということは、すなわち、境界をなす単体を基底として足し算を行った際に、その係数が2の倍数となることとなるとわかる。このようなサイクルの有無をより明示的に与えるため

に、上式の整数値をとる係数 (Z) を、その係数を2で割った際の剰余 (Z_2) で置き換える。これによって先程の端点の足し上げは $2|p_2|+2|p_3|+2|p_4| \rightarrow 0|p_2|+0|p_3|+0|p_4|=0$ となり、境界を足し上げてその和が0となることから、「サイクル」であることの条件と言えるようになった。また、 k 単体 $|a_0 a_1 \cdots a_k|$ から境界を抽出する操作は一般に

$$\partial_k |a_0 a_1 \cdots a_k| := \sum_{i=1}^k |a_0 a_1 \cdots a_{i-1} a_{i+1} \cdots a_k|$$

と記述でき、このような操作は「境界作用素」と呼称される。

我々は「サイクル」を抽出できるようになったが、まだ目標の「穴 (ホモロジー群)」にはたどりついていない。前述のように、1次元の方体複体の部分集合 $\{|p_5 p_6|, |p_6 p_9|, |p_9 p_8|, |p_8 p_5|\}$ や1次元単体複体の部分集合 $\{|p_3 p_4|, |p_4 p_5|, |p_5 p_3|\}$ で囲まれた領域のように、サイクルではあるが、その中身が詰まっている場合があるからである。これには、ターゲットとなるサイクルの1つ上の次元の単体複体 $\{|p_3 p_4 p_5|\}$ を確認して、ターゲットのサイクルがこの高次の単体複体の境界となっていないことを確認すればよいのであった。境界の抽出は境界作用素によって実現されるので、これを高次の単体複体 $|p_3 p_4 p_5|$ に作用させると、

$$\partial_2 |p_3 p_4 p_5| = |p_4 p_5| + |p_3 p_5| + |p_3 p_4|$$

となる。これより、サイクル $\{|p_3 p_4|, |p_4 p_5|, |p_5 p_3|\}$ は高次の単体複体 $\{|p_3 p_4 p_5|\}$ の境界となっており、穴とはならないことが判明する。このようにして、我々は「サイクル」であるが高次の単体複体の「境界」ではないものとして定量的に穴を算出できるようになった。

方体複体では、境界作用素は次のように定義される。

$$\begin{aligned} \partial_k ([I_1, J_1] \times [I_2, J_2] \cdots [I_k, J_k]) \\ := \sum_{i=1}^k [I_1, J_1] \times \cdots \times ([J_i + [I_i]] \times \cdots \times [I_k, J_k]) \end{aligned}$$

例えばこれを、2次元方体に適用すれば、 $\partial_1 ([I_1, J_1] \times [I_2, J_2]) = ([J_1] + [I_1]) \times [I_2, J_2] + ([J_2] + [I_2]) \times [I_1, J_1]$ のように2方体の境界である1方体 (辺) が抽出されることが確認できる。この境界作用素を用いて単体複体と同様の議論を行うと方体複体においても穴の算出ができる。

ここまでは、ある単体複体の部分集合が「穴」になるか否かを調べてきた。最後に、与えられた構造全体に対応する方体/単体複体 $K_{\text{cubical}}, K_{\text{simplicial}}$ からどのようにして個別の「穴」を抽出するかを説明する。**Fig. 1 (c)** 下段の1次元の単体複体に対して境界作用素 ∂_1 を適用することは、複体に含まれる1次元単体を基底とするベクトル空間 $(|p_2 p_3|, |p_3 p_4|, |p_3 p_5|, |p_3 p_6|, |p_4 p_2|, |p_4 p_5|)$ から0次元の単体を基底とするベクトル空間 $(|p_1|, |p_2|, |p_3|, |p_4|, |p_5|, |p_6|)$ への写像

$$\begin{aligned} \partial_1 (|p_2 p_3|, |p_3 p_4|, |p_3 p_5|, |p_3 p_6|, |p_4 p_2|, |p_4 p_5|)^T \\ = (|p_2| + |p_3|, |p_3| + |p_4|, |p_3| + |p_5|, |p_3| + |p_6|, |p_4| + |p_2|, \\ |p_4| + |p_5|)^T \end{aligned}$$

として、**Fig. 1 (d)** 上段のように行列表現できる。ちなみに行列 ∂_1 は一般には正方行列とはならない。この行列を適切に基底変換して、穴に対応するサイクルを要素にもつ基底

から、その境界への写像行列へと変換できたとすると、写像先が0になる写像元の基底がサイクルとわかる。その適切な基底変換を見つける操作は、行列 ∂_k に対して基本変形:

- ・2つの行を入れ替える
- ・2つの列を入れ替える
- ・ある行に他の行を加える
- ・ある列に他の列を加える

を繰り返し適用することによって、行列の (i, j) 成分 (ただし、 $i \neq j$) を0とするような変換: $S_k^{-1} \partial_k T_k = M_k$ を見つけることで実現される。変換によって得られた行列 M_k の (i, i) 成分は1か0の値を取る。このような行列 M_k はスミス標準形と呼ばれる。このようにして得られた基底変換 S_k, T_k が施された1次元単体のベクトル空間

$$(|p_2 p_3|, |p_2 p_4|, |p_3 p_4|, |p_3 p_5|, |p_3 p_6|, |p_4 p_6|) S_k$$

の要素のうち、 M_k の (i, i) 成分が0となる写像元の要素は、0単体の空間で0に写像されるわけなので、この要素が、境界をもたないこと、つまりサイクルになることを意味する。ちなみに、このような操作は写像の核 (Ker) を見つける操作となる。また、 (i, i) 成分が1となる写像元の基底は、写像先の0単体の基底

$$(|p_1|, |p_2|, |p_3|, |p_4|, |p_5|, |p_6|) T_k$$

の対応する線形和を境界としてもつこともわかる。このように、境界作用素をスミス標準形にすることで、単体複体からサイクルと境界を抽出することが可能になる。したがって、1次元の単体複体を作る穴を見つけたければ、1次元単体複体の境界作用素のスミス標準形の (i, i) 成分が0となる要素からサイクルを、2次元単体複体の境界作用素から、1次元の境界作用素で見つけたサイクルが基底となるスミス標準形をみつけ、その (i, i) 成分が1となるか0となるかによって、穴を抽出することが可能となる (**Fig. 1 (d)**)。スミス標準形への変換計算の具体例は、参考文献^{4), 5)}等を参照されたい。

[補遺] 位相的データ分析の実際の計算アルゴリズムでも、この説明と同様のホモロジー群が算出される。一方で、実はこの Z_2 を係数とする算出法ではメビウスの輪のようなねじれ構造の特徴を捕まえることはできない。それを実現するためには、「図形の向き」を導入する必要がある。もし自分の分析対象にねじれ構造の存在が期待される場合、図形の向きを考慮したホモロジー群の算出を検討する必要がある。

2.3 パーシステントホモロジー

物理データの分析においては、密度分布関数を二値化する際の閾値の違いによるトポロジカルな構造の変化 (**Fig. 2** 上段) や、その構造が持つ空間スケール (**Fig. 2** 下段) といった情報が、対象を理解する上で有用である場合がある。このような情報とトポロジカルな特徴量を同時に抽出可能なのが、パーシステントホモロジー群である。パーシステントホモロジー群では、材料構造のパターンを閉凸集合で被服する際に、ある1パラメータ t を導入する。このパラ

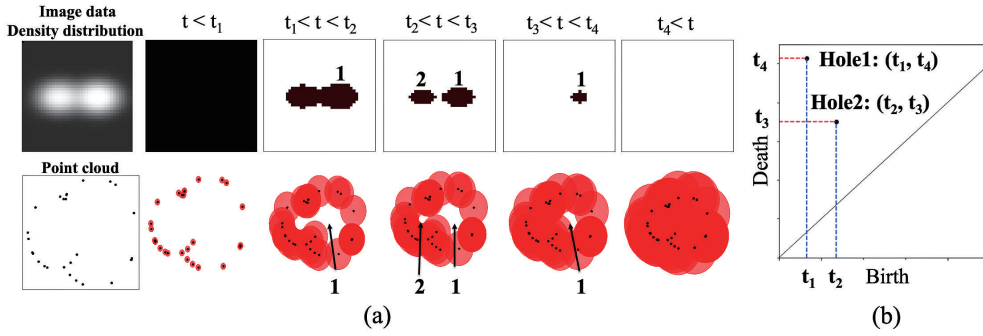


Fig. 2 Filtration and its persistence diagram. In the density distribution, the areas below the threshold are defined as domains.

メータは、例えば、密度分布関数の二値ピクセル画像（立体複体）化であれば、密度関数の閾値に対応し（Fig. 2 上段）、円による点群データの被覆であれば円の半径に対応する（Fig. 2 下段）。このパラメータによる図形の変化を抽出するのがパーシステントホモロジー群である。ただし、パーシステントホモロジー群として扱えるパラメータ t には、その増大とともに必ず単体複体 K_t の要素が増大する $K_t \subset K_{t+}$ という条件が課される。

説明の簡単化のために、パラメータ t の増大を時刻の増大と表現する。その上で、Fig. 2 において図形の 0, 1 次元の穴が新たに生じたり消えたりする特徴的な時刻をそれぞれ t_1, t_2, t_3, t_4 とおく。Fig. 2 の上段はピクセルデータ、下段は点群データにおいて、特徴的な大きな 2 個の穴 1, 2 が生成消滅する例を示している。 t_0 からはじまり時刻の増大に応じて、時刻 t_1 で大きな穴 1 が発生する。時刻 t_2 ではその穴が分裂し、穴 2 ができる。時刻 t_3 では穴 2 が消滅し、時刻 t_4 で全ての穴がなくなる。パーシステントホモロジー群では、このような穴のパラメータに依存した変化を穴の発生時刻を x 軸に、消滅時刻を y 軸にマップすることで表現することが多い（Fig. 2 (b)）。このようなグラフは Persistence Diagram (PD) と呼ばれる。

穴の生成消滅の情報を持つパーシステントホモロジー群も、ホモロジー群と同様に線形代数的操作によって算出することができる。そのためまず、各時刻での単体複体を抽出した上で、それぞれの時刻の単体複体を K_t と設定する（Fig. 3）。次に、時間変化を考慮した上で図形を代数化することを考える。ある穴の時間発展を考える場合、 $K_t \subset K_{t+}$ という関係にあることを仮定したことから、穴の消滅は必ず、その穴を境界とする高次の単体複体の誕生により発生する。従って、時刻に関する情報としては、単体の発生時刻を保持すれば十分であるとわかる。そこで、離散化された時刻の数だけの要素を持ち、その単体や単体の集合（サイクル等）が発生した時刻に、その単体・単体の集合を要素とする基底空間を設定する。例えば、Fig. 3 において t_3 で発生する 2 次元単体 $|p_2 p_3 p_4|$ は以下のように表現される。

$$\langle p_2 p_3 p_4 \rangle = (0, 0, 0, |p_2 p_3 p_4|, 0)$$

ここで $\langle \cdot \rangle$ は、その単体・単体の集合を誕生時刻を表す

基底空間で表現したことを意味するものとした。さらに、この基底に作用すると時刻を発展させるような演算子 x を導入する。つまり、さきほどの 2 次元単体 $\langle p_2 p_3 p_4 \rangle$ に x を作用させると、

$$x(\langle p_2 p_3 p_4 \rangle) = (0, 0, 0, 0, |p_2 p_3 p_4|)$$

となる。これらを使うと、例えば Fig. 3 において t_1 で発生するサイクル $|p_1 p_2| + |p_1 p_3| + |p_2 p_4| + |p_3 p_4|$ は、次のように表現できる。

$$\begin{aligned} & (0, |p_1 p_2| + |p_1 p_3| + |p_2 p_4| + |p_3 p_4|, 0, 0, 0) \\ &= x(|p_1 p_2|, 0, 0, 0, 0) + x(|p_1 p_3|, 0, 0, 0, 0) \\ &+ (0, |p_2 p_4|, 0, 0, 0) + (0, |p_3 p_4|, 0, 0, 0) \\ &= x\langle p_1 p_2 \rangle + x\langle p_1 p_3 \rangle + \langle p_2 p_4 \rangle + \langle p_3 p_4 \rangle \end{aligned}$$

このサイクルの境界は、

$$\begin{aligned} & (0, (|p_1| + |p_2|) + (|p_1| + |p_3|) + (|p_2| + |p_4|) + (|p_3| + |p_4|), 0, 0, 0) \\ &= x(\langle p_1 \rangle + \langle p_2 \rangle) + x(\langle p_1 \rangle + \langle p_3 \rangle) \\ &+ (x\langle p_2 \rangle + x\langle p_4 \rangle) + (x\langle p_3 \rangle + x\langle p_4 \rangle) \\ &= 2x(\langle p_1 \rangle + \langle p_2 \rangle + \langle p_3 \rangle + \langle p_4 \rangle) \end{aligned}$$

となる。このように、ホモロジー群の算出と同様に、パーシステントホモロジー群では、 x の多項式の係数が 2 で割って割り切れる際に、過不足なく境界が重なりサイクルが形成されるとわかる。よって、係数を 2 で割った際の剰余 (Z_2) で置き換えることで、 $2x(\langle p_1 \rangle + \langle p_2 \rangle + \langle p_3 \rangle + \langle p_4 \rangle) \rightarrow 0x(\langle p_1 \rangle + \langle p_2 \rangle + \langle p_3 \rangle + \langle p_4 \rangle) = 0$ のように、足して 0 となることが「サイクル」であることの条件となる。またこの表現から、サイクルは $|p_1|, |p_2|, |p_3|, |p_4|$ の誕生から 1step 後の、 t_1 が誕生時刻とわかる。

次に、算出されたサイクルが消える時刻である、高次の単体の境界になる時刻を算出する方法を説明する。ここまでの議論に基づくと、2 次元単体複体 $|p_1 p_2 p_3| + |p_2 p_3 p_4|$ の境界は以下で与えられる。

$$\begin{aligned} & (0, 0, 0, 0, |p_1 p_2 p_3| + |p_2 p_3 p_4|) \\ &= \langle p_1 p_2 p_3 \rangle + x\langle p_2 p_3 p_4 \rangle \\ &= x^4\langle p_1 p_2 \rangle + x^4\langle p_1 p_3 \rangle + x^2\langle p_2 p_3 \rangle \\ &+ x^2\langle p_2 p_3 \rangle + x^3\langle p_2 p_4 \rangle + x^3\langle p_3 p_4 \rangle \\ &= x^3(x\langle p_1 p_2 \rangle + x\langle p_1 p_3 \rangle + \langle p_2 p_4 \rangle + \langle p_3 p_4 \rangle) \end{aligned}$$

これは、 t_1 で発生したサイクル $x\langle p_1 p_2 \rangle + x\langle p_1 p_3 \rangle + \langle p_2 p_4 \rangle + \langle p_3 p_4 \rangle$ が時刻 $t_{1+3}=t_4$ で 2 次元単体の境界となることを示

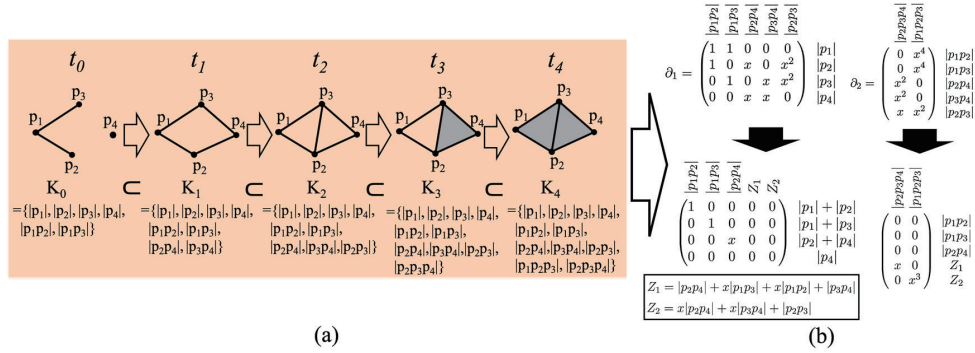


Fig. 3 Procedure of calculation of persistent homology.

している。ここまでの議論に基づくと、このような計算における境界作用素は、一般に次で与えれば良いことが容易に想像されるだろう。

$$\partial_k \langle a_0 a_1 \cdots a_k \rangle := \sum_{i=1}^k x^{b(|a_0 a_1 \cdots a_k|) - b(|a_0 a_1 \cdots a_{i-1} a_{i+1} \cdots a_k|)} \langle a_0 a_1 \cdots a_{i-1} a_{i+1} \cdots a_k \rangle$$

ここで、 $b(\cdot)$ は、その単体が生じた時刻を返す関数であるとした。

ホモロジー群の算出のときと同様に、単体の基底 $\langle \cdot \rangle$ を要素とするベクトル空間を考える。例えば Fig. 3 (a) における 1 単体の基底空間は、以下のように表現される。

$$\langle \langle p_1 p_2 \rangle, \langle p_1 p_3 \rangle, \langle p_2 p_3 \rangle, \langle p_2 p_4 \rangle, \langle p_3 p_4 \rangle \rangle$$

ここで、 $\langle \cdot \rangle$ 自体がその単体の誕生時刻にのみ要素を持つベクトルとなっていることに注意されたい。このようなベクトル空間の元で、パーシステントホモロジー群の境界作用素 ∂_k に対して、ホモロジー群のときと同様に、基本変換を繰り返し適用し、スミス標準形に変形する (Fig. 3 (b))。その際の変換された基底のうち、 (i, i) 成分が 0 となる要素としてサイクルが立ち現れる。また、高次の境界作用素の (i, i) 成分が 1 となる要素と比較することで、それが穴であるかが決定される。ちなみに、ここで標準変形と呼ぶものは、ホモロジー群と違い、「ある行 (列) に他の行 (列) の x^L 倍 (L : 整数) を加える」という変形が含まれる。方体複体に対しても、ホモロジー群の境界作用素を元にパーシステントホモロジー群の境界作用素を導入することで、同様の議論ができる⁸⁾。

このように算出が可能なパーシステントホモロジー群であるが、実際のパーシステントホモロジー群の算出には、全ての次元の単体複体の寿命を同時に算出するような、より洗練された手法が用いられる⁹⁻¹¹⁾。また、そのようなパーシステントホモロジー群の算出アルゴリズムは、複数の位相的数据分析ライブラリから提供されている。その中で我々は、HomCloud¹²⁾ と GUDHI¹³⁾ ライブラリを用いて研究を行っている。どちらも Python から使用可能なライブラリであり、HomCloud は後述する逆解析法が実装された数少ないライブラリの一つであり、Gudhi は後述する PD のベクトル化手法が網羅的に実装され、かつ開発コミュニティと各種ドキュメントが充実したライブラリである。

【補遺】 パーシステントホモロジー群の一意性について

ここまでの議論で、一つの穴が二つに分化した際 (例えば Fig. 3 の t_2)、どちらの穴が元の大きな穴の寿命を引き継ぐかによって PD が変わるのではないかと疑念が生じるかもしれない。実は、基本変換で行列を変換していると、サイクルの表現として複数の基底変換の候補が得られる場合がある。しかしながら、どの基底変換の元でもそのスミス標準形は変わらない。つまり、最終的な穴の寿命の情報は変わらないのである。これは、スミス標準形の一意性として理解されるが、詳細は参考文献^{5), 9)}などを参照されたい。また、参考文献⁵⁾の p. 115 の計算を追うと、このことが体感される。

2.4 なぜパーシステントホモロジーか？

このように本稿の前半分を使って丁寧に説明してきたパーシステントホモロジー群であるが、基本的には穴の生成消滅の情報を抽出しているに過ぎない。したがって、「数学的に厳密と言うだけで、穴を抽出したいだけなら他にも様々な既存方法があるのではないか？」という疑問が生じるかもしれない。本節では、この疑問に対する著者なりの解答を記述したい。

パーシステントホモロジー群の抽出によるデータ分析は、他のトポロジカルな特徴量を用いる分析手法と比較して、以下のような 3 つの長所を「同時に」もつ。

1. 安定性定理¹⁴⁾が成立する
2. 単なるトポロジーではなく空間スケールの情報も適度に含んでいる
3. 逆解析が可能である

1 は、入力構造の微小な変動に対して、PD も微小にしか変動しないことを保証する。このことは、PD が構造の劇的な変化によって特徴付けられる、相転移や核生成の特徴量としての必要条件を満たすことを意味する。つまり、相転移的な激的な構造の変化が起こらない場合に、PD も劇的に変化せず、誤った相転移を検出しないことが保証される。2 は、物理学において重要となる空間スケール情報を PD が含むことを言っており、これが重要であることはすでに述べた通りである。そして 3 目が、パーシステントホモロジーの特に重要な利点である。逆解析とは、構造デー

タから得られた PD から、逆に構造空間上で対応する穴の位置を調べたり、ある PD を再現するような元の構造を逆推定する手法のことである。これによって我々は、この後の機械学習による分析などで得られた知見を、元の構造空間に戻って「解釈」することが可能になる。これは、物理データを分析する上でとても重要な利点である。逆解析の詳細については、参考文献¹⁵⁾を参照されたい。

3. 機械学習による位相的データ分析

機械学習を用いると、PD から物性的、力学的特性を推定することや、その物理的機序を解明することが可能になる^{2), 16), 17)}。本節では、ある構造データが与えられた場合に、位相的データ分析と機械学習を用いてそこから情報を抽出する一連の手順を説明する。分析過程は、大きく分けて「入力データ形式の決定」「PD の算出」「ベクトル化」「機械学習による予測・分類」の4段階で構成される (Fig. 4)。

「入力データ形式の決定」では、与えられた入力データのどのような構造を穴として捉え、そこに存在する何を1パラメータとして設定し、穴の状態の変化をみるかを決定する。例えばグレースケールのピクセル画像データを例に説明すると、前節で説明したように、ピクセルを正方形の凸閉集合として扱い、二値化の閾値を1パラメータとする方法がある。他にも、適当な閾値を決めて2値化した上で、得られたドメイン構造をマンハッタン距離で近いピクセルから順順に膨らませて穴の変化をみる方法、同じく2値化した上で、ピクセルの中心に点群があると考え、そこに半径 r の円を置いて、その r を1パラメータとする方法などが考えられる。

「PD の算出」では、「入力データ形式の決定」で決めたターゲット構造に基づいて、適切な PD 算出アルゴリズムを選択する。例えば、グレースケール画像の場合は、ある閾値 h 以下の領域 (sub level-set) を方体複体とし、 h の増大に対する PD を算出したり⁸⁾、点群データにおいて、点を中心とする半径 r の領域をアルファ複体とし、 r の増大に対する PD を算出する⁵⁾といったことを決定する。

「ベクトル化」について説明する。PD をトポロジカルな特徴量と述べてきたが、実は PD 自体が点群データになっている。これは、光電子分光におけるフォトンカウントデータを分析する際に、ヒストグラム化する必要があるように、

前処理をして初めて分析が可能になることを意味する。機械学習による分析のためには、よく2次元空間上に分布する点群データである PD をベクトルデータに変換する。例えば Persistence image¹⁸⁾ というベクトル化法では、カーネル密度推定を用いて、PD の2次元密度分布関数を算出した上で、グリッド状の座標上の密度分布の値を取得し、それを PD を表現するベクトルデータとする。その際、寿命の短い穴は、元の構造のノイズを反映したものであると考えられるため、ベクトル化する際に、低寿命領域の密度を低くするバイアスをつける場合がある。カーネル密度推定のバンド幅や、前述のバイアスの付け方などはハイパーパラメータとなるため、物性値などのなんらかの既知の値の予測精度を指標として、最適化を行う必要がある。その他にも多数あるベクトル化手法は、Gudhi ライブラリのモジュールである sklearn-tda の web ページ¹⁹⁾に、各手法の論文へのリンクとともにまとめられているため、参照されたい。

「機械学習による予測・分類」では、得られたベクトルデータを用いて物性値・力学特性値の予測を行ったり、構造の分類による相図作成などを行う。さらに、スパースモデリングを用いて、予測や分類にとって重要であった PD の重要構造を推定し、前述した逆推定方を用いることで、元の構造の重要な穴の位置を推定して物理的機序を解明することもよく行われる。最近は、逆推定で得られた重要穴の位置分布をさらに機械学習で分析するといった研究も行われている²⁰⁾。

〔補遺〕分析手順の選択について

このように、位相的データ分析と機械学習を組み合わせることで予測モデルや分類モデルを構築するには、多数の手順と手法の選択、各手法のハイパーパラメータ最適化を行う必要がある。これらの組み合わせによって、分析結果が変わる可能性がある。どの組み合わせが良いかは、予測したい物性値や力学特性値の予測精度が最も良い組み合わせを選んだり、シミュレーションデータであれば、シミュレーション時のモデルパラメータの推定精度を指標として用いることで決定できる。しかし、莫大な数の全ての組み合わせを検証することは難しい。これを回避するには、そもそもデータのどの構造が有用であるかや、得られた PD の分布のどのような構造が重要かといった仮説を立てて、それらの情報抽出に最も有用な手法の組み合わせを選択するという、

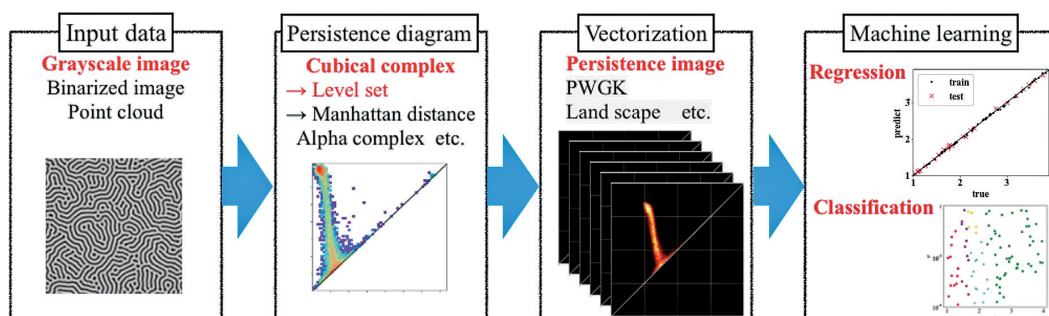


Fig. 4 Analysis procedure of PD using machine learning.

科学者の考察が非常に有用であると著者は考える。

4. 材料構造形成過程の位相的データ分析

4.1 材料構造とパーシステントホモロジー

強磁性体の磁区構造や、構造材料の組織構造、化学反応系の濃度分布構造などの多くのパターン形成過程の計測では、系の第一原理的な素過程における最小単位である、スピンや原子、分子そのものが計測されることは少なく、それらの微小領域における平均値や平均密度・濃度比などが計測されることが多い。このような平均値の空間分布データは、グレースケール画像とみなすことができる。また、そのような系のモデル化においても、連続体近似によって粗視化された場を構成単位として、Time Dependent Ginzburg Landau Equation (TDGL)^{21), 22)} やフェーズフィールドモデル^{23), 24)}、反応拡散モデル²⁵⁾ といった場の時間発展を記述するモデルが用いられることが多い。

グレースケール画像として得られる構造は、よくドメインの繋がり方や、ドメイン数といったトポロジカルな特徴量を用いて分析される²⁶⁻²⁸⁾。これらの分析は、グレースケール画像を2値化した構造に対して適用される。一方で、連続値としての場の分布構造には、ドメイン内のゆらぎなどの情報が埋め込まれている可能性がある。実際に、グレースケール画像の2値化の閾値を変えた際のトポロジカルな構造の変化の様相が、ストライプや亀甲模様のような既知の周期的パターン相の秩序変数のように振る舞うことが報告されている²⁹⁾。材料構造ではしばしば、空間的に非一様な非周期的な秩序構造が現れ、その分類と理解が材料特性やその発生機序を知る上で重要となることがある。もし閾値の変化におうじたトポロジカルな構造の変化を表す特徴量を、このようなパターンの秩序変数として用いることができれば、工学・理学的に有用であると考えられる。一方で、未知のパターン相を検出しそれを理解するためには、得られた特徴量の振る舞いの物理的解釈ができる必要がある。2.4節で述べたように、グレースケール画像の閾値を1-パラメータとしたパーシステントホモロジーは、秩序変数として必要な性質である安定性定理を満たし、物理機序の解明に有用な逆解析が可能のため、このような分析における特徴量として最適であると考えられる。

磁石のような磁性材料において、原子スピンの揃った領域が作る空間的なパターンである磁区構造は、系が保存する磁気的エネルギーの発生機序と密接に関係する重要な特徴である³⁰⁾。例えば、クリップに磁石を近づけると、クリップが磁性を帯びるといった現象もこの磁区構造と関連している。磁性材料の結晶がもつ欠陥、不純物、ステップやひずみといった空間的非一様性に起因した、磁気的配位の非周期的秩序構造は、この磁気的エネルギーの保存の一因と考えられている^{31), 32)}。例えば、空間的非一様性が大きな2次元薄膜上の磁性材料の磁区構造には、その薄膜上に迷路のような非周期的な秩序構造をもつものがある。この迷路において3つの壁が接

地する箇所は、3つの異なる方位特性をもつ磁区の接合部分となるため、3つの壁が等間隔の角度差で配置されない限り部分的に対称性が破れ、そこに磁気的なエネルギーが蓄積される³¹⁾。このように、磁区構造の「かたち」が磁気的特性に本質的に重要であるにもかかわらず、磁区構造の「かたち」と物性特性が具体的にどのような関係性をもつかは、非線形効果が大きな非平衡系の現象であることから、未だに厳密には解明されたとは言い難い状況にある³⁰⁾。したがって、位相的データ分析によって、このような磁区構造を適切に分析できれば、未だ未解明な部分が多い磁区構造と磁性特性の関係の一端が解明されることが期待される。

本研究の目的は、グレースケール画像として得られる磁区構造をパーシステントホモロジーを用いた位相的データ分析法で分析することで、これまで2値化によって失われた情報の中から新たな物理的知見を発見することである。このようなグレースケール画像の分析は、前節で説明したピクセルデータを方体複体として表現した上で、2値化の閾値を1-パラメータとし、閾値以下となる領域(黒ドメイン)を穴とする分析法に対応する。具体的には、TDGLによって数値的に再現された^{21), 22)} 強磁性体の磁区構造の形成過程を分析対象とした。

4.2 TDGL を用いた強磁性体磁区構造の再現

イットリウム鉄ガーネット (YIG) 薄膜上のスピンは、薄膜の垂直方向に強い軸異方性をもつ。そのため、YIG 薄膜上の磁区構造を再現するモデルとして、格子状に配置された微小領域が薄膜の垂直方向に持つスピンの成分の平均値を場 $\phi(\mathbf{r})$, $\mathbf{r}=(x, y)$ とした TDGL が用いられる場合がある。このモデルでは、原子一つ一つのスピンのパターン形成を扱うのではなく、それらの集合の平均場 $\phi(\mathbf{r})$ で構成されるエネルギー関数の最小化としてパターン形成を記述する。つまり、YIG 薄膜上の磁区構造の時間発展を TDGL でモデル化することは、連続体近似によるパターン形成過程のモデル化と理解される。本研究では、工藤らの研究に従い²²⁾、非平衡状態となるような早い速度で減衰する外部磁場中での、YIG 薄膜の磁区構造形成過程を分析対象として、次のような時間発展モデルを設定した。

$$\frac{\partial \phi}{\partial t} = \frac{\delta(H_{\text{ani}} + H_J + H_{\text{di}} + H_{\text{ex}})}{\delta \phi(\mathbf{r})}$$

ここで、 $H_{\text{ani}} + H_J + H_{\text{di}} + H_{\text{ex}}$ は系のエネルギーモデルである。それぞれの項について説明を行う。

H_{ani} は、スピンの上向き ($\phi(\mathbf{r})=1$) または下向き ($\phi(\mathbf{r})=-1$) にそろうこと(異方化)を促すような異方性エネルギーであり、次式で定義される。

$$H_{\text{ani}} = \alpha \int d\mathbf{r} \lambda(\mathbf{r}) \left(-\frac{\phi(\mathbf{r})^2}{2} + \frac{\phi(\mathbf{r})^4}{4} \right)$$

$$\lambda(\mathbf{r}) = 1 + \frac{\mu(\mathbf{r})}{4}$$

ここで、 $\mu(\mathbf{r})$ は平均0分散 σ^2 の正規分布に従う乱数であ

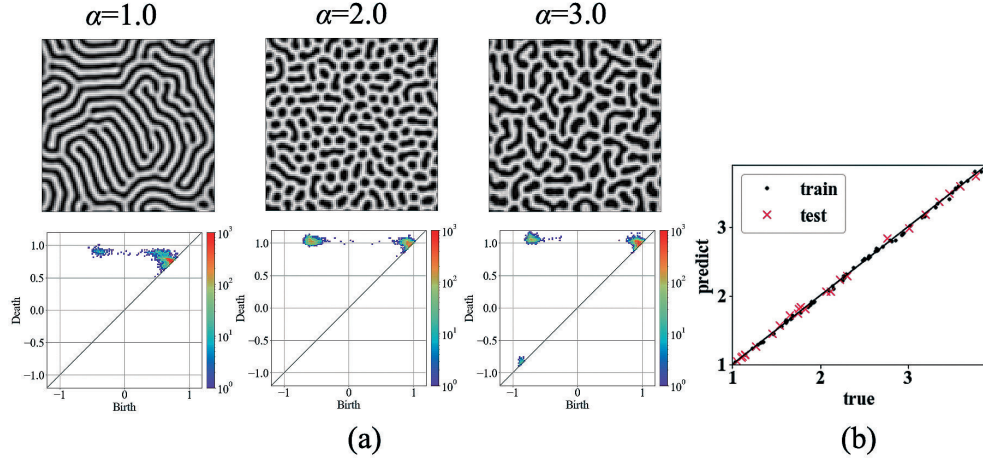


Fig. 5 PD of magnetic domain patterns and regression result.

る。また、 $\lambda(\mathbf{r})$ は、結晶構造の欠陥や不純物の存在などに起因した、磁性材料の空間的な不均質性を現している。磁気異方性の強さは α によってコントロールされており、特に他のエネルギー項と H_{ani} がバランスするような領域では、系の不均一性 $\lambda(\mathbf{r})$ の効果が大きく現れると考えられる。

H_J は $\phi(\mathbf{r})$ のとなりあった微小領域が滑らかに繋がるようにする交換相互作用エネルギーであり、 H_{ani} で異方化されている場合には、同じ値を持つドメインを作る駆動力となる。

$$H_J = \beta \int d\mathbf{r} \frac{|\nabla \phi(\mathbf{r})|}{2}$$

H_{di} は $\phi(\mathbf{r})$ が遠隔領域と異なる値をとるような効果を与える双極子相互作用エネルギーであり、異なる値を持つドメインが組み合わさった構造を作る駆動力となる。

$$H_{\text{di}} = \gamma \int d\mathbf{r} d\mathbf{r}' \phi(\mathbf{r}) \phi(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') \\ G(\mathbf{r}, \mathbf{r}') = \frac{1}{|\mathbf{r} - \mathbf{r}'|^3}$$

H_{ex} は外部磁場の効果を表すエネルギー項で、次で与えられる。

$$H_{\text{ex}} = -h(t) \int d\mathbf{r} \phi(\mathbf{r})$$

外場は時間に依存した関数 $h(t)$ となっており、本研究では、1.5から一定速度 v で減衰し、0.0となるところで変化がなくなるものとした。

TDGLはモデルパラメータ $\alpha, \sigma, \beta, \gamma, v$ によって形成されるパターンを大きく変える。本研究ではその中でも α と v に着目した。これは、 α が磁性材料の不均一性という物性的特性と関連していること、及び、 v が非平衡過程と密接に関連したパラメータであるためである。実際に、この α と v によって異なる磁区パターンと同様のパターンが実材料の測定から観測されることも報告されている²²⁾。このような磁区構造の分析では、 $\phi(\mathbf{r})$ を2値化した上で、得られたドメイン構造の孤立ドメイン数やドメイン長さといった特徴量をみる分析が行われる。したがって、前述したように、連続値をとる場 $\phi(\mathbf{r})$ を連続値のまま扱いつつ分

析することで、これまでに得られていない物理的な知見を得られる可能性がある。

本研究では、このTDGLのシミュレーションで得られたパターン形成過程のデータに対して位相的データ分析を適用することで、材料構造の分析でよく行われる、「物性値との回帰」「磁区構造の分類（相図作成）」「パターン形成ダイナミクスの分析」を行なった。それぞれについて次節で説明を行う。

5. 結果と考察

α と v を変化させながらTDGLに基づいて時間発展の数値実験を行い、外部磁場が0となつてから十分に時間がたった状態としてFig. 5 (a) 上段のような磁区構造が得られた。本章では、得られた磁区構造に位相的データ分析を適用した結果について説明する。そこでまず磁区構造のPDとそのベクトル化データの算出手順をまとめる。まず、格子状に配置された $\phi(\mathbf{r})$ がピクセルデータとして解釈できることから、磁区構造を立方体複体として取り扱った。また、 $\phi(\mathbf{r})$ は連続値をとる平均磁化の分布関数であることから、グレースケール画像と同様に、閾値を1-パラメータとしてパーシステントホモロジー群を算出した。特に本研究では、磁化がマイナスとなる領域（グレースケールでの黒い領域）を穴と定義したため、負の分布関数 $-\phi(\mathbf{r})$ の閾値以下の領域に立方体があるとして分析を行なった。この結果、Fig. 5 (a)にあるような、 α の値によって分布構造が変化するPDが得られた。

次に、ベクトル化には、前述したPersistent image³³⁾を用いた。Persistent imageでPDの密度分布を得るために用いたカーネルには、ガウスカーネルを用い、そのカーネル幅とベクトル化する際のグリッドサイズというハイパーパラメータは、後述する物性値との二乗予測誤差を最も最小化するように、交差検証法を用いて決定された。このベクトルデータは、磁区構造の分類による相図作成の際にもそのまま用いられた。これは、その現象にとって重要と考えられる物性値をよく予測するハイパーパラメータは、その対象の情報をよく抽出できているだろうという考えに基づく。

5.1 物性値との回帰

本研究では物性値として $\phi(\mathbf{r})$ の磁気異方性の強さや、磁性材料の空間的不均一性の強さを表す α を、系の特性と強く関連した予測すべき物性値 (のようなもの) として設定した。その上で、磁区構造を位相的データ分析で特徴量化し、それを用いて α を推定することを試みた。推定モデルには線形回帰モデルを用いた。ベクトル化された PD の次元数は 10,000 である一方で、磁区構造データは 100 サンプル程度しかないため、高次元空間中にあるシンプルなモデルを、少ないデータ数の元で効率的に抽出するスパースモデリング法の一つである Least absolute Shrinkage and Selection Operator (LASSO)³⁴⁾ を用いた。

回帰分析の結果、Fig. 5 (b) にあるような予測モデルが構築された。Fig. 5 (b) の横軸は TDGL によるデータ生成時に設定された α の正解値であり、縦軸は PD を用いた回帰モデルによる予測値である。また、表中の丸点は訓練データを、× 点は回帰モデルの訓練に用いなかったテストデータでの予測結果を表している。図からわかるように、高い精度の予測モデルが構築されており、1 から 4 の値を取る α のレンジに対して、0.03 程度のばらつき (二乗平均平方根誤差) での予測が実現された。また、分析の詳細は頁数の関係で割愛するが、PD の逆解析の結果、ドメインの形の情報だけでなく、ドメイン内の $\phi(\mathbf{r})$ の強度揺らぎの情報を含めることで、この高い回帰精度が実現されていることも判明した。この分析結果は、まさに 2 値化せずに $\phi(\mathbf{r})$ を連続値として扱うことで得られた結果と言えよう。また、現状の磁区構造の測定手法に、このドメイン内ゆらぎを検出する手法はないが、この分析結果は、ドメイン内ゆらぎを検出する有用性を示唆している。

5.2 磁区構造の分類

Fig. 5 (a) 上段にあるように、 α と v によって磁区構造は大きく構造を変える。工藤らは、この違いを結晶化における過飽和度の違いと対応づけ、過飽和度の増大に応じて発生ドメイン数が増大するという視点から、迷路構造 (Fig. 5 (a) 上段の $\alpha=1.0$) と島構造 (Fig. 5 (a) 上段の $\alpha=2.0$) という分類を行なった²⁶⁾。本研究では、パーシステントホモロジー群という視点から、この分類をさらに精緻化することを試みた。

4.1 節の回帰分析時に得られた PD のベクトル化データに対して、階層的クラスタリング法³⁵⁾ と呼ばれる分類方法を適用した。この手法は、ベクトルの近さに応じてデータを分類する手法であり、ある 2 つの集団を異なるクラスタとみなす際の、集団間の距離の閾値によって分類の細かさを変えることができる。この分析の結果、ある分類距離の閾値において、Fig. 6 のような分類結果が得られた。Fig. 6 (a) はシミュレーション時に用いたパラメータ α の値を横軸に、 v の値を縦軸にとり、図中の色点はそれぞれ、別のシミュレーションデータに対応している。また、色の違いは分類分析で得られたクラスタを、I から VI の数字は、Fig. 6 (b)

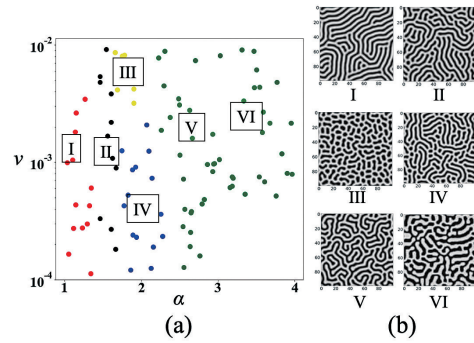


Fig. 6 Classification of magnetic domain pattern.

の各ドメインパターンが含まれるクラスタを表している。Fig. 6 (b) は、外部磁場が 0 になった後、十分な時間経過後に得られる磁区構造である。

先行研究で島構造と定義されていた領域は Fig. 6 (a) の III に対応し、迷路構造は V 及び VI に対応する。一方で、Fig. 6 (b) をみると、I、II の領域も迷路構造となっていること、及び IV は島と迷路が混ざったような混合相となっていることがわかる。つまり、 α の増大に応じて、「迷路 → 島 (混合相) → 迷路」と変化するのである。これまでの研究で行われていたような孤立ドメイン数による分類では、I と V、VI の迷路構造を区別することはできないが、パーシステントホモロジーを用いた分析では、それらが分類されるのである。さらに、詳細は割愛するが、それぞれの迷路クラスタの中も、さらに分類可能であることもわかった。また、混合相 IV の発見も位相的データ分析による新たな発見と言えよう。次節のダイナミクスの分析から、このような不思議な相図が得られた背景の一端も判明した。

5.3 パターン形成ダイナミクスの分析

ここまでは、外部磁場が 0 になってから十分に時間が経過後に形成される最終的な磁区構造についての分析であった。本節では、その磁区構造が形成されるまでの時間発展を位相的データ分析法を用いて分析する。分析は、異なる時刻の磁区構造に対して個別に PD を算出し、その PD の時間変化を分析することで行われた。その結果、Fig. 7 のような結果が得られた。Fig. 7 (a) は、 α が小さな領域 (I, II) で迷路ができる時間発展を分析した結果であり、Fig. 7 (b) は、 α が大きな領域 (V, VI) で迷路構造ができる時間発展を分析した結果である。また、図中にある “Pattern Formation” 行は、磁区構造そのものの時間発展を、“Histogram” 行は、強度分布である $\phi(\mathbf{r})$ の強度ヒストグラムの時間発展を、“PD” 行は PD の時間発展を、“Fourier” 行は磁区構造を 2 次元フーリエ変換した結果の時間発展をそれぞれに表している。特に Fig. 7 (a)、(b) のそれぞれの 2 列目と 3 列目は、磁区構造のパターン核の生成時刻と、パターン核が成長して迷路構造を作る時刻をそれぞれ取り出している。PD 行をみると、同じ磁区構造でも、その核生成と核成長の状態が Fig. 7 の (a) と (b) で大きく異なることが読み

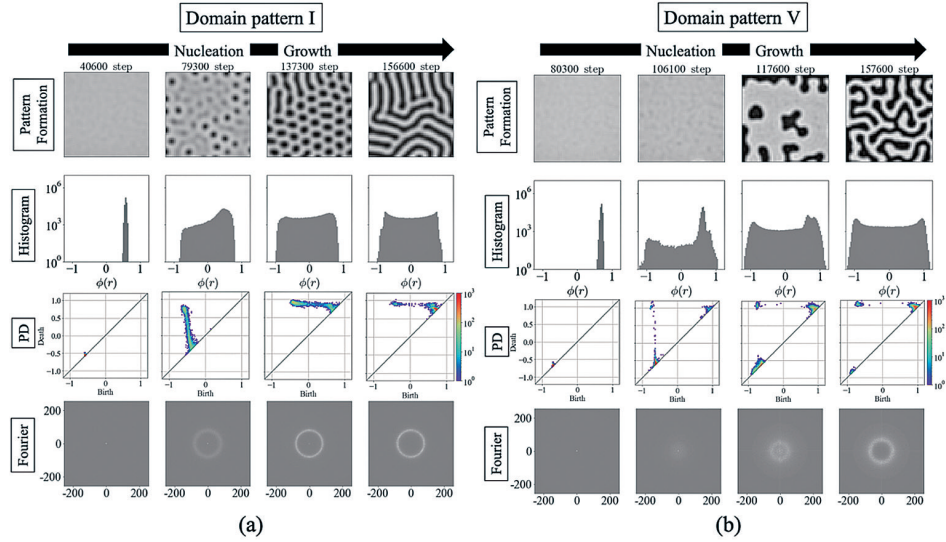


Fig. 7 Time evolution of magnetic domain pattern and its features.

取れる。つまり、(a) はつながった細長い分布構造を持つ一方で、(b) はおおよそ3点のみに分布を持つことがわかる。一方で、“Histogram”や“Fourier”からは、このような明瞭な違いは読み取れない。

PDの逆解析の結果から、(a)の細長い構造は、 $\phi(r)$ が白ドメインでも黒ドメインでもない、中間の値を取ることを示していることがわかった。つまり、 α が小さな領域では、 $\phi(r)$ の強度の意味で中間状態を経てドメイン構造を形成し、 α が大きな領域では、一気に $\phi(r)$ が反転しドメイン構造を形成するのである。これは、 α によって強度がコントロールされる H_{ani} の標式に戻って考えれば理解される。つまり、 α は磁気異方性の強度を表しているため、この値が小さいと中間状態をとりやすく、大きいととりにくくなるのである。これが、パターン形成過程に影響を与えたと考えられるのである。一方で、同じ迷路構造も異なる過程で構成可能とみることができる。これは、工学的に磁区構造をコントロールする必要性が生じた際に、有用な知見になる可能性がある。また、 $\alpha=2.0$ 近辺でPDは中間状態を取らなくなることも判明した。これは、島構造IIIや混合相IVの形成においても、この中間状態を取れるかどうか重要な要因となっていることを示唆している。これらもまさに、 $\phi(r)$ を連続値として扱うことによって得られた知見と言えよう。

6. まとめと今後の展望

1章で述べたように、結晶構造に対してはフーリエ基底が、熱力学のようなランダムな現象に対しては統計量空間が有用である。一方で、本研究のような非線形・非平衡なパターン形成過程では、Fig. 7からわかるように、PDがその振る舞いを最もよく捉えられるとわかった。これは、パーシステントホモロジー群に基づく特徴量抽出が、非周期的な秩序構造を形成するパターン形成過程の理解とモデル化を行うための特徴量空間として有効であることを示唆す

る。対象とする系や、1-パラメータの設定方法が異なるものの、本研究以外にも、このような仮説を担保する位相的データ分析の適用事例が多数報告されている^{2), 3), 16), 17)}。

本研究から判明したように、PD空間でのパターン形成過程のモデル化を考える上で、Fig. 7(b)のように分布が不連続に変化する構造を捉えることが重要となる。我々はこのような特徴的な構造をもつPD空間のダイナミクスのモデル構築について、2つのアプローチを考えている。一つは、本研究のように各時刻で独立に算出されたPDの時間発展を、分布の飛びも含めてモデル化が可能な隠れマルコフモデル³⁶⁾によってモデル化することである。もう一つは、時間発展全体を表現するPDを算出し、そのパラメータ依存性を推定するモデルを作るアプローチである。

後者のアプローチは、時間発展全体をそのまま一つの特徴量として扱う野心的なアプローチであるが、その実現には大きな困難が予想される。例えば、物理時間を1-パラメータとしてパーシステントホモロジーを考えることは困難である。なぜならば、前述したようにパーシステントホモロジーの1-パラメータには、その増加に応じて、単体複体が増大列をなしていなければならないからである。仮に、この問題が生じないように物理系を限定しても、我々はパターンの時空間構造をモデル化したいわけなので、時間と空間の2-パラメータによるパーシステントホモロジーを考える必要が生じる。しかし、2-パラメータのパーシステントホモロジーを考えるには、スミス標準形は一変数多項式でしか定義できないので、複数パラメータのPHはこの方針では定義できないという問題が立ち塞がる。実際に、パーシステントホモロジーを2-パラメータに拡張することが数学的に非常に困難であるとされている。また、時間と空間を合わせた時空間での距離を1-パラメータとすることも考えられるが、時間と空間をどのような比率で取り扱うかという問題が生じる。このように、トポロジー構造の時間

発展の分析は、位相的データ分析におけるこれからの課題であると思われる。一方で、応用者の立場から、数学的な厳密さを一時的に離れ、いくつかの直感に基づく時間発展データへの手法拡張を行い、物理現象の予測・理解への効果の大きさを持って、数学側に新たな視点を与えるというアプローチの研究の推進も期待される。

謝 辞

本研究は、JSPS 科研費 新学術領域研究「次世代物質探索のための離散幾何学」20H04648、及び JST CREST JPMJCR15D3、JPMJCR1861、及び統計数理研究所共同プログラム (2021-ISMCRP-2069) の助成を受けたものです。本原稿作成にあたっては、統計数理研究所の栗木哲教授には多数の有益な助言を頂きました。また、産業技術総合研究所の中村壮伸主任研究員との議論から、物理データ分析における PH の長所についての深い洞察を得ました。両名に心より感謝いたします。

参 考 文 献

- 1) C. Gunnar: "Topology and data", *Bulletin of the American Mathematical Society*, **46**-2 (2009), 255-308.
- 2) Y. Hiraoka, T. Nakamura and A. Hirat: "Hierarchical structures of amorphous solids characterized by persistent homology", *Proceedings of the National Academy of Sciences*, **113**-26 (2016), 7035-7040.
- 3) 大林一平: "位相的データ解析の現在(統計的モデリングと予測理論のための統合的数理研究)", 数理解析研究所講究録 = RIMS Kokyuroku, 2057 (2017), 34-50.
- 4) K. Fukumizu: Machine Learning Approach to Topological Data Analysis, <https://www.ism.ac.jp/~fukumizu/courses.html>, Intensive Course at Tokyo Metropolitan University, (2018).
- 5) 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群入門, 共立出版, (2013).
- 6) C. Kittel, M. Paul and M. Paul: Introduction to solid state physics. 8, New York: Wiley, (1996).
- 7) K. Masao, I. Obayashi, Y. Takeichi, R. Murao and Y. Hiraoka: "Non-empirical identification of trigger sites in heterogeneous processes using persistent homology", *Scientific reports*, **8**-1 (2018), 1-9.
- 8) T. Kaczynski, M. Konstantin and M. Marian: Computational homology. Springer Science & Business Media, 157 (2006).
- 9) M. Buchet, Y. Hiraoka and I. Obayashi: "Persistent homology and materials informatics", *Nanoinformatics*. Springer, 75-95, Singapore, (2018).
- 10) H. Edelsbrunner, D. Letscher and A. Zomorodian: "Topological Persistence and Simplification", *Proceedings 41st annual symposium on foundations of computer science. IEEE*, (2000), 454-463.
- 11) N. Otter, M. A. Porter, U. Tillmann, P. Grindrod and H. A. Harrington: "A roadmap for the computation of persistent homology", *EPJ Data Science*, **6**-1 (2017).
- 12) 大林一平: "パーシステントホモロジーに基づくデータ解析パッケージ HomCloud の紹介 (第 16 回 生物数学の理論とその応用 - 生命現象の定量的理解に向けて -)", 数理解析研究所講究録 = RIMS Kokyuroku, 2166 (2020), 103-107. (HomCloud URL: "https://homcloud.dev/index.html")
- 13) Dłotko, Paweł: "Computational and applied topology, tutorial." arXiv preprint arXiv: 1807.08607 (2018). (GUDHI URL: "https://gudhi.inria.fr")
- 14) D. Cohen-Steiner, H. Edelsbrunner and J. Harer: "Stability of persistence diagrams", *Discrete & computational geometry*, **37**-1 (2007).
- 15) 大林一平: "パーシステンス図の逆問題", *応用数理*, **26**-4 (2016), 151-158.
- 16) T. Ichinomiya, I. Obayashi and Y. Hiraoka: "Persistent homology analysis of craze formation", *Phys. Rev. E*, **95**-1 (2017), 012504.
- 17) 山田拓洋, 鈴木雄太, 三俣千春, 小野寛太, 上野哲朗, 大林一平, 平岡裕章, 小嗣真人: "パーシステントホモロジーを用いた迷路状磁区構造におけるトポロジカル欠陥の可視化," *表面と真空*, **62**-3 (2019), 153-160.
- 18) H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta and L. Ziegelmeier: "Persistence images: A stable vector representation of persistent homology", *The Journal of Machine Learning Research* **18**-1 (2017), 218-252.
- 19) Saul. Nathaniel and Tralie. Chris.: "Scikit-TDA: Topological Data Analysis for Python. Zenodo." (2019) (sklearn-tda URL: <https://github.com/MathieuCarriere/sklearn-tda>)
- 20) T. Ichinomiya, I. Obayashi and Y. Hiraoka: "Protein folding analysis using features obtained by persistent homology", *Biophysical Journal*, **118**-12 (2020), 2926-2937.
- 21) E. A. Jagla: "Numerical simulations of two-dimensional magnetic domain patterns", *Physical Review E*, **70**-4 (2004), 046204.
- 22) K. Kudo and M. Mino: "Magnetic domain patterns depending on the sweeping rate of magnetic fields", *Journal of the Physical Society of Japan*, **76**-1 (2007), 013002-013002.
- 23) 小山敏幸, 小坂井孝生, 宮崎亨: "計算機シミュレーションによる材料組織の予測と解析 Phase-field 法による相分解組織形成の計算機シミュレーション", *まてりあ*, **38**-8 (1999), 624-628.
- 24) C. Long-Qing: "Phase-field models for microstructure evolution", *Annual review of materials research*, **32**-1 (2002), 113-140.
- 25) P. Gray and S. K. Scott: "Autocatalytic reactions in the isothermal, continuous stirred tank reactor: isolas and other forms of multistability", *Chemical Engineering Science*, **38**-1 (1983), 29-43.
- 26) K. Kudo and K. Nakamura: "Field sweep-rate dependence of magnetic domain patterns: Numerical simulations for a simple Ising-like model", *Physical Review B*, **76**-5 (2007), 054111.
- 27) J. Guiu-Souto, J. Carballido-Landeira and P. M. Alberto: "Characterizing topological transitions in a Turing-pattern-forming reaction-diffusion system", *Physical Review E*, **85**-5 (2012), 056205.
- 28) Y. Nahas, S. Prokhorenko, Q. Zhang, V. Govinden, N. Valanoor and L. Bellaiche: "Topology and control of self-assembled domain patterns in low-dimensional ferroelectrics", *Nature communications*, **11**-1 (2020), 1-8.
- 29) K. R. Mecke: "Morphological characterization of patterns in reaction-diffusion systems", *Physical Review E*, **53**-5 (1996), 4794.
- 30) A. Hubert and R. Schäfer: "Magnetic domains: the analysis of magnetic microstructures", Springer Science & Business Media, (2008).
- 31) M. Seul, L. R. Monar and L. O'Gorman: Pattern analysis of magnetic stripe domains morphology and topological defects in the disordered state, *Philosophical Magazine B*, **66**-4 (1992), 471-506.

- 32) N. Abu-Libdeh and D. Venus: “Dynamics of topological defects in a two-dimensional magnetic domain stripe pattern”, *Physical Review B*, **84**-9 (2011), 094428.
- 33) H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta and L. Ziegelmeier: “Persistence images: A stable vector representation of persistent homology”, *The Journal of Machine Learning Research* **18**-1 (2017), 218-252.
- 34) R. Tibshirani: “Regression shrinkage and selection via the lasso”, *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**-1 (1996), 267-288.
- 35) L. Rokach and O. Maimon: “Clustering methods”, *Data mining and knowledge discovery handbook*, Springer, US, (2005), 321-352.
- 36) L. E. Baum and T. Petrie, “Statistical Inference for Probabilistic Functions of Finite State Markov Chains”, *The Annals of Mathematical Statistics*, **37**-6 (1966), 1554–1563.

代表者メールアドレス

本武 陽一 mototake@ism.ac.jp