

位相データ解析について

徳永 健一*

A Review of Topological Data Analysis from the View of Industrial Application

Ken-ichi Tokunaga*

本稿では、近年産業界における応用面でも注目されている位相データ解析について、これから適用を検討している人を対象にその概要を紹介する。また、後半ではあまり知られていない、ある位相データ解析の計算方法を紹介する。

Keywords: 数値解析、位相データ解析、ホモロジー、絡み数

1. はじめに

位相データ解析とは、位相幾何学（トポロジー）の手法を用いたデータ解析手法のことであり、インターネットおよび計算科学の発展により規模が大きくなったデータに対して、その「形」に関する特徴量を取り出すものとして、近年注目されている。ソーシャルネットワーク、金融、センサーなどのビックデータと呼ばれるようなものや、医療、材料科学などの分野にも応用されている。

本稿では、位相データ解析で現在もっともよく用いられているパーシステントホモロジーについて概説し、次に積分公式によって位相不変量を計算する方法をいくつか紹介する。

2. パーシステントホモロジー

2.1. ホモロジー

トポロジーでは、連続写像で移りあう図形同士は同じもの（同相）とみなす。同相かどうかを区別するために、同相であれば同じ値になる不変量（位相不変量）を計算する。ホモロジー群は位相不変量の一つである。直感的には図形の「穴」を計算するものである。0 次のホモロジーは連結成分の数、1 次のホモロジーはリングの数、2 次のホモロジーは空洞の数に対応する。

初等的な多面体の場合、頂点を 0 次元、辺を 1

*アドバンスソフト株式会社 第 2 事業部

次元、三角形を 2 次元として、 i 次元の図形を基底とするベクトル空間 C_i （より一般には生成する加群）を考え、その上で、 i 次元の図形から $i-1$ 次元の境界を対応させるような写像 $\partial_i: C_i \rightarrow C_{i-1}$ （境界写像）を定義する。このときホモロジー群は $H_i = \text{Ker} \partial_i / \text{Im} \partial_{i+1}$ で定義される。

一般には、ベクトル空間の基底となる図形（から定義されるもの）を単体と呼び、それを用いて定義されるベクトル空間と境界写像の組 $\{(C_i, \partial_i)\}$ のことを複体と呼ぶ。ホモロジーとは複体から図形の「穴」を代数的に計算する手法の一種である。

複体を構成する方法は多面体から初等的に構成する方法のほかに、特異複体、チェック (Cech) 複体、アルファ複体などがある。

2.2. パーシステントホモロジー

パーシステントホモロジーは複体のフィルトレーション（複体の間の写像が定義された複体の列）から計算する。

点群から複体のフィルトレーションを作る例を紹介する。それぞれの点を中心とする半径 r の閉球を考える。閉球それぞれを 0 単体、2 つ繋がった閉球の組を 1 単体、3 つの閉球がそれぞれつながった 3 つ組を 2 単体、のように定めて、境界を定める演算子を定義すると、複体になる。厳密には Cech 複体と呼ばれるものである。詳細は専門書 [1] を参照していただきたい。なお、数値計算ではアルファ複体が使われることが多い。

この時、半径 r を大きくすると複体 $C(r)$ の間には自然な包含写像が定義できる。従って、半径を大きくしていったときにできる複体の列がフィルトレーションになる。

$$C(r_1) \hookrightarrow C(r_2) \hookrightarrow \cdots \hookrightarrow C(r_n)$$

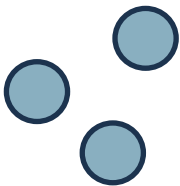
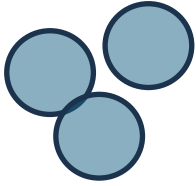
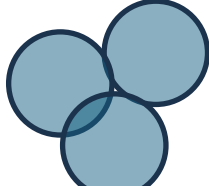
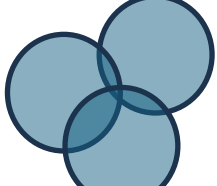
単純な点群の場合の例を表 1 に示す。

フィルトレーションに対してホモロジーを計算したものがパーシステントホモロジーである。

$$H_l(r_1) \rightarrow H_l(r_2) \rightarrow \cdots \rightarrow H_l(r_n)$$

半径を大きくするにつれて、途中で「穴」が発生し、さらに大きくすると「穴」が消滅する。対応するホモロジー群において、この「穴」に対応するホモロジー類を得ることができる。ホモロジー類が $r = b$ で発生し、 $r = d$ で消滅するものを $I(b, d)$ で表すと、パーシステントホモロジーはこれらの直和で記述することができる（一意分解定理）。横軸に b 縦軸に d 、をとってパーシステントホモロジーの元をプロットしたものがパーシステント図と呼ばれるものである。

表 1 複体のフィルトレーション

			
$C_0 = \{x_0, x_1, x_2\}$	$C_0 = \{x_0, x_1, x_2\}$ $C_1 = \{x_{01}\}$	$C_0 = \{x_0, x_1, x_2\}$ $C_1 = \{x_{01}, x_{12}, x_{20}\}$	$C_0 = \{x_0, x_1, x_2\}$ $C_1 = \{x_{01}, x_{12}, x_{20}\}$ $C_2 = \{x_{012}\}$
	$\partial x_{01} = x_0 - x_1$	$\partial x_{01} = x_0 - x_1$ $\partial x_{12} = x_1 - x_2$ $\partial x_{20} = x_2 - x_0$	$\partial x_{01} = x_0 - x_1$ $\partial x_{12} = x_1 - x_2$ $\partial x_{20} = x_2 - x_0$ ∂x_{012} $= x_{01} + x_{12} + x_{20}$
$H_0 = 3K$	$H_0 = 2K$	$H_0 = K$ $H_1 = K$	$H_0 = K$

3. 位相不変量の積分公式

多様体上の積分によって、その多様体の位相不変量を求める方法をいくつか紹介する。

2.3. パーシステントホモロジーの応用

パーシステントホモロジーの応用例として最も知られているものとして、シリカガラスの原子配置に対するパーシステントホモロジーを計算して、結晶状態とも液体状態とも異なる特徴を見出した報告 [2]がある。ガラスのパーシステント図には特徴的な曲線が現れ、シリカ原子のリング構造の幾何学的な制約と考えられる。

その他、タンパク質構造やセンサーネットワーク、ソーシャルネットワークにパーシステントホモロジーを適用された例が報告されている。

2.4. パーシステントホモロジーを計算するソフトウェア

日本で開発されている HomCloud [3]の他、GUDHI [4]、Dipha [5]、Dionysus2 [6]、JavaPlex [7]等が知られている。

3.1. ガウス・ボンネの定理

積分で位相不変量を求める最も典型的で古くから知られている例として、ガウス・ボンネの定

理がある。向き付け可能なコンパクト2次元リーマン多様体 M 上のガウス曲率 K の積分とオイラー数 $\chi(M)$ の間には次の関係が成り立つ。

$$\int_M K \, dV = 2\pi\chi(M) \quad (1)$$

これは局所的な量である曲率から大域的な量であるオイラー数が求められるという意味で、ミクロとマクロをつなぐ公式であるといえる。

3.2. 絡み数

もう一つ別の積分で定義できる位相不変量として、絡み数を紹介する。3次元空間内の2つの有向閉曲線があるとき、一方がもう一方の周りをどれだけ回っているかを表す整数値を絡み数という。絡み数が0ではないもっとも単純な例はホップ絡み目である。ホップ絡み目の絡み数は1である。

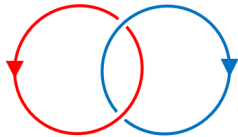


図1 ホップ絡み目

絡み数は位相不変量の一つである。これは以下のように2つの有効閉曲線を J および K とすると、積分で与えられる。

$$L(J, K) = \frac{1}{4\pi} \int_J \int_K \frac{y - x}{|y - x|^3} \cdot dx \times dy \quad (2)$$

3.3. 折れ線の絡み目

3次元空間内の有向閉曲線を折れ線で近似して、折れ線の頂点の列 $J = \{x_0, x_1, x_2, \dots, x_{n-1}, x_n\}$ 、 $K = \{y_0, y_1, y_2, \dots, y_{m-1}, y_m\}$ で表す。ここで $x_0 = x_n$ および $y_0 = y_m$ とする。この時、折れ線それぞれで $x = x_i + (x_{i+1} - x_i)s$ および $y = y_j + (y_{j+1} - y_j)t$ のように変数変換して

$$\int_0^1 \int_0^1 \frac{|y_j - x_i, x_{i+1} - x_i, y_{j+1} - y_j| \, ds \, dt}{|(y_j - x_i) + (y_{j+1} - y_j)t - (x_{i+1} - x_i)s|^3} \quad (3)$$

を L_{ij} とすると、積分公式(2)は

$$L(J, K) = \frac{1}{4\pi} \sum_{i=1}^n \sum_{j=1}^m L_{ij} \quad (4)$$

となる。(3)式を変形して $u_0 = y_j - x_i$

, $u_1 = x_i - x_{i+1}$, $u_2 = y_{j+1} - y_j$ とおけば、

$$L_{ij} = - \int_0^1 \int_0^1 \frac{|u_0, u_1, u_2| \, ds \, dt}{|u_0 + u_1 s + u_2 t|^3} \quad (5)$$

と変形できる。これは $Q_0 = u_0$, $Q_1 = u_0 + u_1$, $Q_2 = u_0 + u_1 + u_2$, $Q_3 = u_0 + u_3$ としたとき、三角形 $Q_0 Q_1 Q_2$ を原点から見たときの立体角の値 $\Omega(Q_0, Q_1, Q_2)$ を用いて、

$$L_{ij} = -\Omega(Q_0, Q_1, Q_2) - \Omega(Q_0, Q_2, Q_3) \quad (6)$$

と表すことができる。折れ線の頂点で表すと $Q_0 = y_j - x_i$, $Q_1 = y_j - x_{i+1}$, $Q_2 = y_{j+1} - x_{i+1}$, $Q_3 = y_{j+1} - x_i$ となるので、以上をまとめると

$$\begin{aligned} L(J, K) = & -\frac{1}{4\pi} \sum_{i,j}^{n,m} \left(\Omega(y_j - x_i, y_j - x_{i+1}, y_{j+1} - x_{i+1}) \right. \\ & \left. + \Omega(y_j - x_i, y_{j+1} - x_{i+1}, y_{j+1} - x_i) \right) \end{aligned}$$

となる。

3.4. 立体角の公式

三角形 $Q_0 Q_1 Q_2$ を原点から見たときの立体角の値は次の公式(7)で与えられる[8]。

これらを合わせると、有効閉曲線を折れ線で近似すれば、絡み数は逆三角関数の計算で求められることがわかる[9][10]。

3.5. ホップ絡み目三角形の計算

三角形のホップ絡み目の絡み数を上記の公式を用いて計算する例を示す。これは三角形同士が自己交差するかどうかを判定する式にもなっているので、三角形ポリゴンの処理やメッシュ生成処理においても用いることができる。ただし計算量は従来の方法よりも多くなるので注意していただきたい。

三角形の頂点の座標を

$$\begin{aligned} x_0 &= (1, 0, 0), \\ x_1 &= (-2, 2, 0), \\ x_2 &= (-2, -2, 0) \\ y_0 &= (-1, 0, 0), \\ y_1 &= (2, 0, -2), \end{aligned}$$

$$y_2 = (2,0,2)$$

とする。ここで $Q_{ij} = y_j - x_i$ として計算した結果を示す。

$$\Omega(Q_0, Q_1, Q_2) = 2 \tan^{-1} \left(\frac{|Q_0, Q_1, Q_2|}{|Q_0||Q_1||Q_2| + (Q_0, Q_1)|Q_2| + (Q_1, Q_2)|Q_0| + (Q_2, Q_0)|Q_1|} \right) \quad (7)$$

i	j	L_{ij}	
0	0	$-\Omega(Q_{00}, Q_{10}, Q_{11})$ $-\Omega(Q_{00}, Q_{11}, Q_{01})$	2.656150981715479
1	0	$-\Omega(Q_{10}, Q_{20}, Q_{21})$ $-\Omega(Q_{10}, Q_{21}, Q_{11})$	0.6867995926646446
2	0	$-\Omega(Q_{20}, Q_{00}, Q_{01})$ $-\Omega(Q_{20}, Q_{01}, Q_{21})$	2.656150981715479
0	1	$-\Omega(Q_{01}, Q_{11}, Q_{12})$ $-\Omega(Q_{01}, Q_{12}, Q_{02})$	0.6867995926646443
1	1	$-\Omega(Q_{11}, Q_{21}, Q_{22})$ $-\Omega(Q_{11}, Q_{22}, Q_{12})$	-0.805431683161323
2	1	$-\Omega(Q_{21}, Q_{01}, Q_{02})$ $-\Omega(Q_{21}, Q_{02}, Q_{22})$	0.6867995926646443
0	2	$-\Omega(Q_{02}, Q_{12}, Q_{10})$ $-\Omega(Q_{02}, Q_{10}, Q_{00})$	2.656150981715479
1	2	$-\Omega(Q_{12}, Q_{22}, Q_{20})$ $-\Omega(Q_{12}, Q_{20}, Q_{10})$	0.6867995926646446
2	2	$-\Omega(Q_{22}, Q_{02}, Q_{00})$ $-\Omega(Q_{22}, Q_{00}, Q_{20})$	2.656150981715479

これらの和はほぼ 4π となり、絡み数が 1 であることが計算できた。

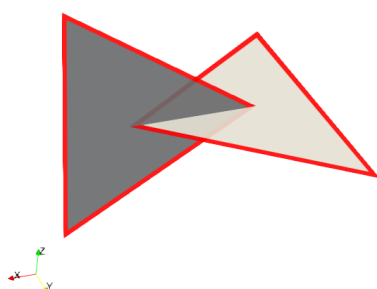


図 2 三角形ホップ絡み目

3.6. 応用例

環状の高分子化合物、DNA 鎖など環状の構造をもつものが絡み合っているかどうかは絡み数を使って判定することができる。

結び目ができている高分子は通常よりも小さくなり、溶液中の移動速度が変化するなどの物性の違いが現れる。

DNA 鎖の絡み数の変化によって部位特異的組み換え酵素の働きを解明することもできる。

4. まとめ

本稿ではパーシステントホモロジーの概要、絡み目数を積分で計算する方法、について紹介した。材料科学などで今後さらに重要になると考えられる。従来の手法では見るのが難しかった特徴量を取り出す手法として興味ある方は是非お試しください。当社はこの手法を用いた解析や開発が対応可能であるので、ご連絡いただきたい。

参考文献

- [1] 服部晶夫, 位相幾何学, 岩波書店, 1991.
- [2] Y. Hiraoka, T. Nakamura, A. Hirata, Y. Nishiura, “Hierarchical structures of amorphous solids characterized by persistent homology,” *Hiraoka, Yasuaki, et al. "Hierarchical structures of amorphous solids characterized bProceedings of the National Academy of Sciences*, 第 卷 113, 第 26, pp. 7035-7040, 2016.
- [3] “HomCloud,” [オンライン]. Available: <https://www.wpi-aimr.tohoku.ac.jp/TDA/homcloud/>.
- [4] [オンライン]. Available: <https://gudhi.inria.fr/>.
- [5] [オンライン]. Available: <https://github.com/DIPHA/dipha/>.

- [6] [オンライン]. Available:
<https://mrzv.org/software/dionysus2/>.
- [7] [オンライン]. Available:
<https://appliedtopology.github.io/javaplex/>.
- [8] A. V. Oosterom and J. Strackee, "The Solid Angle of a Plane Triangle," *IEEE Transactions on Biomedical Engineering*, Vols. BME-30, no. 2, pp. 125-126, 1983.
- [9] Z. Arai, "A rigorous numerical algorithm for computing the linking number of links," *Nonlinear Theory and Its Applications*, 第 巻 4, 第 1, pp. 104-110, 2013.
- [10] M. Bright, O. Anosova, V. Kurlin, "A Formula for the Linking Number in Terms of Isometry Invariants of Straight Line Segments," *Comput. Math. and Math. Phys.*, 第 巻 62, pp. 1217-1233, 2022.

※ 技術情報誌アドバンスシミュレーションは、アドバンスソフト株式会社 ホームページのシミュレーション図書館から、PDF ファイル（カラー版）がダウンロードできます。（ダウンロードしていただくには、アドバンス/シミュレーションフォーラム会員登録が必要です。）