

令和元年6月14日現在

機関番号：13601

研究種目：基盤研究(C) (特設分野研究)

研究期間：2016～2018

課題番号：16KT0131

研究課題名(和文)代数トポロジー的データ解析の地理学への応用に関する研究

研究課題名(英文) Application of topological data analysis to geographic information

研究代表者

沼田 泰英 (Numata, Yasuhide)

信州大学・学術研究院理学系・准教授

研究者番号：00455685

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究では、代数的トポロジー的な解析手法を用いて地理データ(GISデータ)を解析することを主たる目的としている。その解析に必要な代数的トポロジーの理論および技術や、地理学の視点からの問題抽出などについて、周辺分野の研究者も巻き込み議論を行うと共に、実際に地理データに関し計算代数的トポロジー的な解析手法を適用することが目標であった。数理経済談話会というセミナーを行い分野を超えて議論を行うとともに、これらの議論で得た知見を参考に、関連する代数的トポロジーの研究を進め、公共施設の位置情報などについて計算機による代数的トポロジー的解析を試みた。

研究成果の学術的意義や社会的意義

パーシステントホモロジーは代数的トポロジー的データ解析手法の一つであり、材料科学などの主に自然科学の分野で、従来にはできなかったデータ解析を可能にしてきたが、地理データへの応用はなかった。パーシステントホモロジーは、互いに独立した点のデータに外部からネットワークの構造を付加した上で、そのネットワークを解析しているものと思うことができるため、地理データが暗にもつネットワーク構造と強い相関をもつデータに地理データを粗視化することが期待でき、地理データへのパーシステントホモロジーの応用は自然である。

研究成果の概要(英文)：The purpose of this research project was to try to apply methods of topological data analysis to data of geographic information systems. We had some seminars to discuss with researchers in the related areas. In the seminar, we discuss and study the related algebraic topology and methods of topological data analysis. Moreover we calculated the persistence homologies of data of positions of public facilities, e.g., bus stops in some cities. We tried to classify the data sets by differences of persistence homologies.

研究分野：表現論的組合せ論

キーワード：代数的トポロジー パーシステントホモロジー 地理情報

1. 研究開始当初の背景

パーシステントホモロジーなどの代数的トポロジーの理論を用いたデータ解析手法は、計算機の発達などとも相まって近年発達してきた。実際のデータ解析に用いられる中その解析に必要となる基礎理論も発達してきている。応用と基礎理論が研究の両輪として機能し、応用分野には新しい手法が提供され、一方基礎となる数学の分野には新しい問題が提供され、近年盛んに研究されている。

代数トポロジー的データ解析の代表的な手法の一つであるパーシステントホモロジーとは大雑把には次のものである。点の集合として与えられるデータに対し、それらの点を中心とする半径 r の球の和集合を考え、この図形ホモロジーを考える。半径ごとにホモロジーが決まるが、これらのホモロジーの族を Quiver の表現だと思えば直既約分解を行う。このとき得られる直既約成分は、図形に現れる‘穴’がいつ生まれいつ消えるのかを表していると考えることができる。ここでは、半径をパラメータとして考えたが、一般には半径である必要はなく、パラメータで変化する図形であれば、同様のことを考えることができる。

トポロジー的な特徴量は大きさのような情報を持たず、‘穴の個数’や‘連結成分の個数’といった情報しか持たないのが通常である。一方、パーシステントホモロジーは、穴がいつ生まれ消えるかという情報を持っており、‘大きさ’に関する情報を適度に持っているため、より精密な解析ができると思われる。実際、この精密化は功を奏し、材料科学などの分野においてパーシステントホモロジーを用いたデータ解析が応用されている。例えば、液体、結晶、ガラスという各状態における分子配置は、それぞれ特徴的なパーシステントホモロジーを持っているため、パーシステントホモロジーを特徴量として状態を分類するという研究のような応用がある。

一方、温度変化による状態変化を捉えようとするとき、半径というパラメータの他に温度というパラメータも同時に扱う必要がある。パラメータが2つに増えたことにより、Quiver の表現としては難しい対象を扱う必要が出てくる。より具体的には、格子グラフを台とする Quiver の表現は、有限表現型と呼ばれるものではないので、直既約成分として現れる可能性が無限個あり、それらを比較するのが難しい。しかしながら、今解析したい状況の表現は、一般の表現ではなく、必ず図式の可換性という関係式をもつ表現である。したがって、関係式をもつ Quiver の表現を考えればよく、然るべき関係式をもつ特定の Quiver の表現論について表現型を与えよという数学的な問題が生まれる。通常個別の Quiver の表現は動機がないため個々に研究されることは少ないが、データ解析への応用という明確な動機が与えられ、意味のある問題として捉えることができる。実際にこのような動機から、格子を台にもつ可換図式に相当する関係式をもつ Quiver の表現型が研究され、サイズが小さいときには有限型であるということが調べられている。

パーシステントホモロジーを用いたデータ解析手法は、材料科学などの自然科学には盛んに応用されており、研究が進んでいる。しかしながら、地理情報のようなデータへの応用はあまり扱われていない。また、人文科学への応用というのも、まだ進んでいないように思われた。パーシステントホモロジーによるデータ解析自体は、‘点の位置情報’という数値データがあれば可能である。また、‘点の位置情報’というのも実は実際にある点の座標である必要はなく、単なるベクトルデータであればよく、次元も高次元でも構わない。ベクトルの多重集合として捉えられるデータであれば基本的には適用することができる手法であり、非常に汎用な手法であるため、あまり応用の進んでいない分野においても適用することが可能なデータは多数あると思われた。

2. 研究の目的

本研究の目的は、パーシステントホモロジーによるデータ解析手法などの代数的トポロジーの理論を用いたデータ解析手法を、地理データに適用し、解析を試みるとともにその解析により生まれる新しい数学上の問題を考察することにある。

研究の背景の節で説明したように、パーシステントホモロジーによるデータ解析自体はベクトルの多重集合として捉えられるデータであれば適用できる汎用な手法である。しかしながら、材料科学などの諸分野では応用例があるものの、地理学における応用例はないように見受けられた。地理情報は、地図アプリケーションやカーナビゲーションシステムなどの発達により、いわゆる地図情報システム(GIS)のデータとして電子情報として蓄えられているものが多い。また、それらのデータには、オープンデータとして公開されているものも少なくない。その中には公共施設の位置情報なども含まれ、大きなデータセットをなしている。そのため、パーシステントホモロジーなどの手法を用い、データを粗視化し解析することは有益だと思わ

れる。

公共施設などの配置は住民の生活などの行動原理に即して有機的に配置されていると思われる。公共交通機関に関わる施設の場合には、交通ネットワークという形でそのネットワークを認識ができる。交通ネットワークの様なネットワークを明示的に伴った施設は少なく、通常は単独の点が無数に存在しているということ以上の情報はない。しかし、どの種類の施設でも某かの影響を互いに受けながら配置が決まっているものだと思われる。その配置を支配する原理は、交通機関に関わる施設や住民の福祉に関わる施設などそれぞれの性格や種類によって異なると思われる、それらの配置に強く影響を与えていると思われる。また、施設の配置は、その地域の住民生活に関わるものであり土地ごとによって変わってくる可能性もある。特に、その土地の成り立ちなどにも影響を受けているものと思われる。これらの配置を支配する原理の違いは複雑に互いに影響を与えているため、その配置は単純なものではない。

パーシステントホモロジーの代表的な適用方法では、与えられた点の多重集合を中心とした半径 r の円を考える。この際半径である r より近い距離にある点の一つの連結成分になり、もともと単独の点の多重集合でしかなかったデータにネットワークの構造が加わる。考える半径を動かすことでそのネットワークの構造が変わるが、その変化を抽出するのがパーシステントホモロジーである。つまり、互いに独立した点のデータに無理やり外部からネットワークの構造を付加した上で、そのネットワークを解析しているものと思うことができる。この外部から付加したネットワーク構造は、解析対象が明示的もしくは非明示的に持っている、真のネットワーク構造とは異なるはずではあるが、そのネットワーク構造が配置の情報に強い影響を与えているのであれば、外部から付加したネットワークと真のネットワークに大差はないことが期待できる。したがって、パーシステントホモロジーが真のネットワークの特徴量を捉えているということが期待できる。本研究では、地理データにパーシステントホモロジーを用いたデータ解析手法を用いることで、地理データが暗にもつネットワーク構造による地理データの分類を行うことが目的である。

3. 研究の方法

本研究では、代数的トポロジーの理論に基づく解析手法の中でも、パーシステントホモロジーによる解析を主に用いて研究を進める。研究の目的の節で述べたように、パーシステントホモロジーは、互いに独立した点のデータに無理やり外部からネットワークの構造を付加した上で、そのネットワークを解析しているものと思うことができるため、地理データが暗にもつネットワーク構造と強い相関をもつデータに地理データを粗視化することが期待できるためである。地図情報システム(GIS)のデータとして公開されている電子情報をオリジナルのデータとし、公共施設等の配置について解析をする。公共施設の種別や都市名、時期といったメタデータごとにデータセットを切り分け、それらごとにパーシステントホモロジーを計算し、その情報によってデータセットを分類する。その分類はメタデータの情報の分類の影響を受けていることが期待できる。そのパーシステントホモロジーの計算にはオープンソースソフトウェアとして公開されているものを利用し、必要に応じて追加のプログラムを実装する。

本研究は、数学者と地理学者のチームによる共同研究である。実際にデータを解析することによって必要となる新たな理論的な問題については主に数学を専門とするメンバーが、データ解析を行う対象やデータセットの切り分けの方針などの地理学の知見が必要となる問題については主に地理学を専門とするメンバーが担当し、互いに協力をし研究を進める。

また、異なるバックグラウンドをもつ研究者によるチームであるため、研究分野間の意思疎通を円滑にすることが必要である。そのために、ミーティングやセミナーを持つ。とくに、今後関連する可能性のある分野の専門家を招聘し数理経済談話会という形での講演を依頼する。地理データに対する代数的トポロジー的解析手法は前例がないため知見があまりないので、現時点では明白な強い関連はないように見えるものであっても、今後関連する可能性のある分野については積極的に講演を依頼し、関連分野を巻き込む形で研究を進める。

4. 研究成果

本研究では、代数的トポロジー的な解析手法を用いて地理データ(GISデータ)を解析することを主たる目的としている。その解析に必要な代数的トポロジーの理論および技術や、地理学の視点からの問題抽出などについて、周辺分野の研究者も巻き込み議論を行うと共に、実際に地理データに関し計算代数トポロジー的な解析手法を適用することが目標であった。数理経済談話会というセミナーを行い、関連するもしくはこれから関連する周辺分野の研究者を招聘し概説及び研究動向について講演してもらい、関連する話題について議論を行った。扱ったテーマは、応用を視野に入れた統計分野の話題から純粋数学の組合せ論の話題まで幅広く、様々な視点から活発な議論を行うことができた。

これらの議論で得た知見を参考に、国土交通省がインターネット上で公開している地理データ (GIS データ) について、計算機による代数的トポロジー的解析を試みた。特に、施設の位置情報の多重集合ととらえ、パーシステントホモロジーを用いた解析を試みた。特に、対象とする地理的範囲を動的に変更するなどの手法を試行錯誤した。ここで行った解析結果、つまり、得られたパーシステント図に関する情報、および、それらをもとにメタデータを分類した結果については、今後公開予定である。

また、田中は、代数的トポロジーの理論的研究を進めた。より具体的には、単体複体の一般化として有限でサイクルのない小圏と 復体についての理論研究を進めた。特に、ホモトピー論について考察し、strongly collapsibility の特徴付けを与えた。この研究成果については論文 [1] として発表済みである。

5 . 主な発表論文等

〔雑誌論文〕(計 1 件)

[1] K. Tanaka. Strong homotopy types of acyclic categories and n -complexes. Applied Categorical Structures, 査読有, 27(3), 245-260 (2019).
doi: 10.1007/s10485-018-9552-0

6 . 研究組織

(1)研究分担者

研究分担者氏名：武者 忠彦

ローマ字氏名：(MUSHA, Tadahiko)

所属研究機関名：信州大学

部局名：学術研究院社会科学系

職名：准教授

研究者番号 (8 桁): 70432177

研究分担者氏名：田中 康平

ローマ字氏名：(TANAKA, Kohei)

所属研究機関名：信州大学

部局名：学術研究院社会科学系

職名：助教

研究者番号 (8 桁): 70708362

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。