

自然災害予測のための地形情報に基づく位相的データ解析

増田 圭汰¹ 佐藤 好久²

概要: 情報技術の発展に伴い、人間では全体を把握することが難しい巨大なデータ群はこれまで以上に注目を集めている。近年、データ解析技術に対する新しい手法として位相的データ解析 (TDA) が注目されている。これはあるデータに対して、そのデータが持つ「形」を位相的な観点から解析する手法である。本研究では、TDA の有用性を確かめるために、TDA を利用することで地形情報が含まれる画像から特徴抽出を行えるか検証し、また、その結果から自然災害予測へ応用することを目的としている。尚、本研究では対象を水害に限定して行った。本研究を進めていくことで、惑星のような直接現地へ赴くことが困難な未開の地について、地形画像から災害予測が可能になることを期待する。

キーワード: 情報数学, 画像分類, 機械学習

Topographical information-based topological data analysis for natural disaster prediction

Abstract: With the development of information technology, huge groups of data that are difficult for humans to grasp in their entirety are attracting more attention than ever before. In recent years, Topological Data Analysis (TDA) has been attracting attention as a new approach to data analysis techniques. This is an analysis method that extracts the features of a certain data from the topological information of its "shape". In order to confirm the usefulness of TDA, this study aims to verify whether TDA can be used to extract features from images that contain topographical information, and to apply the results to natural disaster prediction. We focus on flood damage. We hope that this research will make it possible to predict disasters based on topographic images for undeveloped areas such as planets, where it is difficult to visit the site directly.

Keywords: Information Mathematics, Image Classification, Machine Learning

1. はじめに

情報技術の発展に伴い、人間では全体を把握することが難しい巨大なデータ群はこれまで以上に注目を集めている。そして、膨大なデータから有益な情報を抽出するデータ解析技術の進化が求められている。近年、データ解析技術に対する新しい手法として位相的データ解析 (TDA: Topological Data Analysis) が注目されている。TDA とは、あるデータに対して、そのデータが持つ「形」を位相的

な観点から解析する手法である。従来の統計的データ解析は、対象となるデータを既知の分布に当てはめて解析をするため、想定した分布が適切でない場合、解析がうまくいかないことがある。TDA はデータの分布モデルに依存しないため、様々なデータに対して解析を行うことができる。

本研究では、TDA の有用性を確かめるために、TDA を利用することで地形情報が含まれる画像 (地図や地形図等) から特徴抽出を行えるか検証し、また、その結果から自然災害予測へ応用することを目的としている。しかし、数ある自然災害を同時に検証することは難しいと判断した。そのため、本研究では水害 (特に洪水) に限定して検証を行っている。

本研究を進めていくことで、惑星のような直接現地へ赴くことが困難な未開の地について、地形画像の情報から災

¹ 九州工業大学大学院 情報工学府 情報創成工学専攻
Department of Creative Informatics, Graduate school of
Computer Science and Systems Engineering, Kyushu Institute
of Technology

² 九州工業大学大学院 情報工学研究院 知能情報工学系
Faculty of Computer Science and Systems Engineering,
Kyushu Institute of Technology

害予測が可能になることを期待する。

2. 位相的データ解析

まず、本研究にて扱う TDA の手法の一つである「パーシステントホモロジー群」について説明していく。

2.1 パーシステントホモロジー群

パーシステントホモロジー群は、図形の連結成分や輪っか、空洞といった構造に注目することで、データの「形」として情報を抽出する手法である。パーシステントホモロジー群を簡単に理解するために、図1のような点群を用いて説明する。

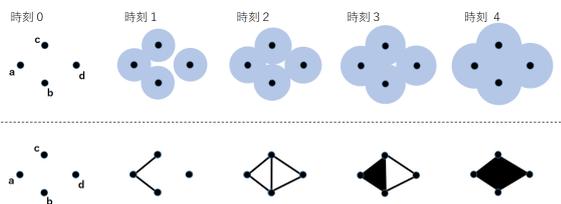


図 1: パーシステントホモロジー群

図1の時刻0において、4つの点が存在するとする。この点群に対し、時間経過と共にそれぞれの点を中心とした球を徐々に大きくしていくことを考える。二つが交わったところを線分、3つが交わったところを中の詰まった三角形とする。時刻1にて、球同士が交わり図中の点a,b,cが線分となり消滅していることがわかる。つまりこの三点は、発生時刻 (birth) が時刻0であり、消滅時刻 (death) が時刻1と表すことができる。また時刻2にて、図中の点a,b,cと点b,c,dがそれぞれ交わり、それぞれ穴を生成している。点a,b,cにおける穴は時刻3にて消滅し、点b,c,dにおける穴は時刻4にて消滅している。従って、点a,b,cにおける穴は発生時刻 (birth) が時刻2であり、消滅時刻 (death) が時刻3と表せる。さらに、点b,c,dにおける穴は発生時刻 (birth) が時刻2であり、消滅時刻 (death) が時刻4と表せる。このように決めた (birth,death) についての集まりをパーシステントダイアグラムといい、0次元パーシステントダイアグラムは点について、1次元パーシステントダイアグラムは三角形などの閉じた折れ線について考える。また、パーシステントダイアグラムを基に、横軸を発生時刻 (birth)、縦軸を消滅時刻 (death) とし、birth,deathの時刻対の集まりを平面上にプロットした図をパーシステント図という。図2は図1における0次元パーシステント図であり、図3は図1における1次元パーシステント図である。

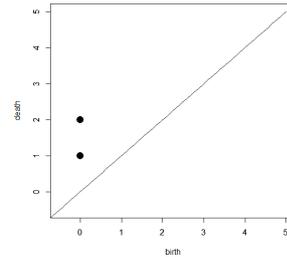


図 2: 図1における0次元パーシステント図

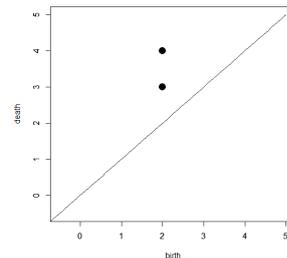


図 3: 図1における1次元パーシステント図

一般的に、パーシステント図における対角線付近に存在する点は、生成元の存在区間が短いため、形の特徴を反映しておらず、ノイズになりやすい。逆に、存在区間が長いものに関しては、情報価値が大きくなる。

2.2 白黒二値画像の解析

白黒二値画像による TDA では、白黒二値画像に「マンハッタン距離」を導入し、その増減に伴う形の変化を調べることでパーシステントダイアグラムを生成する [4], [5], [7].

マンハッタン距離は、あるピクセルから別のピクセルまで到達するために横切る辺の数の最小値で与えられる。黒の領域の内側にも距離は定義でき、それを負の値で表す。黒の領域を1ピクセル太らせる操作は、マンハッタン距離が1の白ピクセルを黒ピクセルにすることに対応する。

例として図4を考える。図4の左側から順番に時刻-3, 時刻-2, 時刻-1, 時刻1とする。図中左側の赤で囲まれた領域と青で囲まれた領域について、発生時刻を-3とする。また、この領域は図の中央の時刻-2において、他の領域とつながることで消滅している。このとき、0次元パーシステント区間は $[-3,-2]$ である。また、図中の時刻-2において、緑色で囲まれた領域 (黒の領域を増やしたことで発生した領域) は、時刻-2にて発生し、図のまま黒の領域を太らせていくと時刻1にて消滅する。このとき、1次元パーシステント区間は $[-2,1]$ である。以上のように、二値化の閾値を徐々に変化させることで、領域の増大列を作ることができ、パーシステントダイアグラムが計算できる。

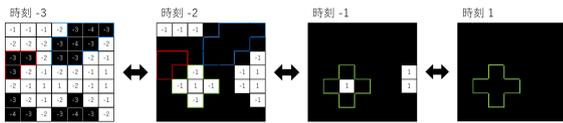


図 4: 白黒二値画像の解析の例

3. 機械学習

大量のデータに対して人間が行うような判断をコンピュータに自動的に実行させるものが機械学習である。本研究では、そのうち「主成分分析」「クラスター分析」「 k -NN法」を利用しており、それらの手法について説明する。

3.1 主成分分析

主成分分析 (PCA: Principal Component Analysis) は、多くの変数により記述された量的データの変数間の相関関係を利用して、情報の損失を最小限に抑えながら、小数の相関のない変数に「まとめて」、まとめられたいくつかの「本質的な」変数（合成変数）によって、データの構造を解析する手法である。

主成分分析では、**主成分**とよばれる合成変数を作成する。これは変数間の重み付き和による合成変数を作成する方法である。このときの重みは、作成される合成変数の分散が最大になるように重みを付ける。このように、分散が最大になるように作られた重みの重み付き和のことを**第1主成分**という。分散が大きくなるような変数を用いれば、その変数の値の大小で違いが明確になると考える。また、第1主成分とは直交するように、つまり相関が生じないように、かつ分散が最大になるように作った重みを利用して作成した重み付き和のことを**第2主成分**という。第1主成分や第2主成分とも直交するように分散を最大にするような重みの重み付き変数を**第3主成分**という。以下同様に重み付き和として合成変数を作成することができ、もとの変数に特に変な関係がない限り、元々あった変数まで主成分を作ることができる [12], [13]。

図5は主成分分析による次元削減の例である。

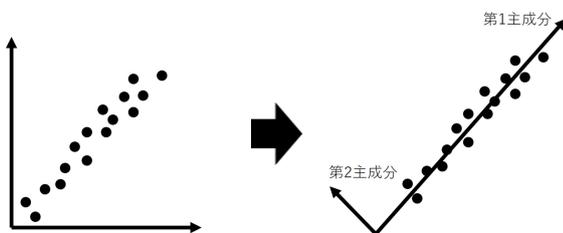


図 5: 主成分分析における次元削減の例

図5において、左側はある2つの変数からなるデータを示している。このデータに対して、図の右側のように分散が最大になるように作られた第1主成分とそれに対する第

2主成分を生成する。ここで、第1主成分のみで上手くデータが説明されていることがわかる。つまり、元は2つの変数で説明されていたデータを1つの変数のみで説明できるようになり、2次元が1次元に削減されている。

3.2 クラスター分析

クラスター分析とは、分類問題を取り扱う教師なし学習である。どの個体（データ）がどのグループに属するかに関する事前情報がない状態（ラベルがないデータ）でグルーピングを行う学習方法である。本研究では「非階層クラスター分析」の手法の一つである k -means法を使う。

図6は k -means法の手順を簡単に表したものである。 k -means法は、いくつかのクラスターに分類するかをあらかじめ決めておき、サンプルを分割していく手法である。まず、 k 個の初期クラスターの中心を適当に与える。その後全てのデータと各クラスターの中心との距離を求め、データを最も近いクラスターに振り分け直し、新しく作成したクラスターの中心を求める。これを結果が収束するか、あらかじめ設定した繰り返し回数の最大値に達するまで繰り返すことで分析を行う [14]。

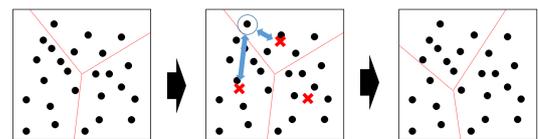


図 6: 非階層クラスター分析の手順

3.3 k -NN法

線形判別関数による判別方法以外の判別分析を非線形判別分析といい、このうち、多数決による判別方法の一つとして k 最近傍法 (k -NN法)がある。この手法は伝統的なパターン分類アルゴリズムであり、判別すべき個体の周辺の個体で最も近い物を k 個見つけ、その k 個の多数決によりどのグループに属するかを判別する方法である。例として図7を挙げる。

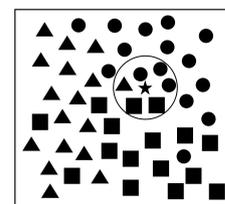


図 7: k -NN法の例

図7は $k = 6$ の場合である。未知データである★を中心に最も近いものを6個探している。図中円内に●が3個、▲が2個、■が1個であることから、未知データ★は●のグループに属すると判断される。

4. 検証方法

4.1 手順

今回扱う画像は「等高線メーカー」サイト [8], [9] と地理院地図 Vector[10] を基に作成している。詳細は節 4.2 にて説明する。作成したデータに対し、いくつかのアプローチにて検証を行っている。また、本研究ではパーシステントホモロジーにもとづいたデータ解析ソフトウェアである Homcloud を用いて TDA を行っている [6], [7].

まず二値画像の地形画像に対しそのまま TDA を行った場合における検証を行う。この検証の流れを図 8 に示す。0次元パーシステントダイアグラムは時刻 0 以降に点が発生しないため分布に偏りがある。そのため今回は 1次元パーシステントダイアグラムを中心に扱っていく。TDA により得られた 1次元パーシステントダイアグラムをベクトル化する。ベクトル化は図 9 のようにパーシステントダイアグラムをメッシュ化し、その枠の内に存在する個数を基に作成する。縦横それぞれ 5 等分にし、対角線から上の情報のみを抜き出すことで 15次元のベクトルに変換している。このとき、パーシステント図の対角線付近に存在する点は生成元の存在区間が短いため、形の特徴を反映しておらず、情報価値が少ないと判断できる。そのため、重み付け ($\arctan(0.01 \cdot l^3)$, $l = \text{death} - \text{birth}$) を行うことにより対角線付近の情報価値を低くし、逆に対角線から離れるほど価値が上がるようにしている。その後、主成分分析を行い、その結果に対して非階層クラスター分析 (k-means 法) を行う。非階層クラスター分析より、地形画像から特徴抽出が行えているかを確認する。

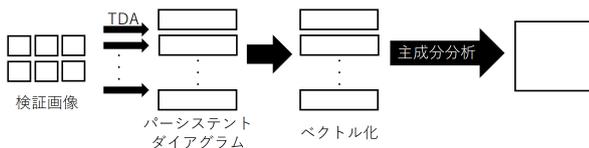


図 8: 研究の流れの概略図 1

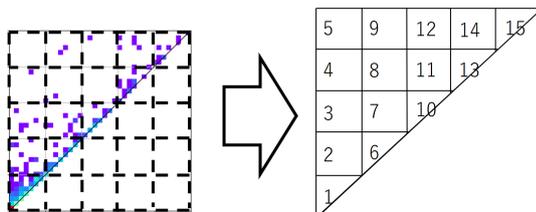


図 9: 1次元 PD のベクトル化

次に二値画像の地形画像を基に生成したポイントクラウドに対し TDA を行った場合における検証を行う。この検証の流れを図 10 に示す。得られたパーシステントダイア

グラムをベクトル化し主成分分析を行う。ベクトル化は、二値画像の地形画像に対しそのまま TDA を行った時と同様、1次元パーシステントダイアグラムを 15次元のベクトルに変換している。また、重み付けも同様に行っている。ベクトル化の範囲を変動させて、特徴抽出の検証に適切な範囲を探る。

また、 k -NN 法を二つのアプローチで行う。一つ目は適切と判断されたベクトル化で得られたデータに対して、 k -NN 法により自然災害予測を行う。このとき、0次元パーシステントダイアグラムも図 11 のように 5次元のベクトルに変換し、1次元とあわせて 20次元のベクトルとしている。二つ目はパーシステントダイアグラムに対する距離ボトルネックディスタンスによる k -NN 法を用いて自然災害予測を行う。パーシステント図の全体集合に対して距離関数を定義することができ、この距離関数を用いることで、二つの異なるデータに TDA を行って得られたパーシステントダイアグラムの比較を行うことができる。この手法は以上の距離関数として、ボトルネックディスタンスを採用したものである。

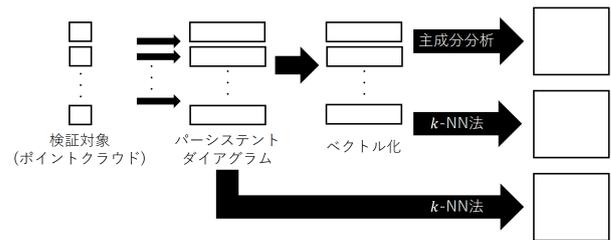


図 10: 研究の流れの概略図 2

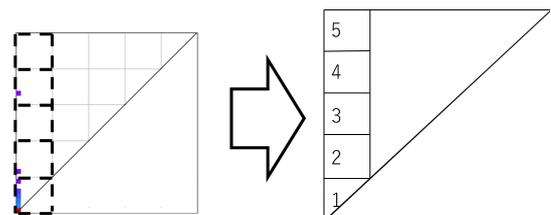


図 11: 0次元 PD のベクトル化

4.2 データ作成

本研究にて使用するデータは、基本的に「等高線メーカー」サイト [8], [9] と地理院地図 Vector[10] を基に作成している。データの作成の流れとしては、まず「等高線メーカー」にて等高線のみを含む地形画像を、地理院地図 Vector[10] にて水域 (川や池等) のみを含む画像をそれぞれ取得する。等高線の取得間隔は 20m としている。両方にて作成した画像を、対応する場所の画像同士で合成することにより、等高線と水域が含まれる地形画像画像を作成する。本研究では、二値画像に対して TDA を行っていく

ため、作成したデータは全て二値画像に変換しており、黒いピクセル部分が等高線と水域である。節 4.1 にて説明した手法 1 の「二値画像の地形画像に対しそのまま TDA を行った場合」では、以上にて作成した画像を使用している。また、手法 2 の「二値画像の地形画像を基に生成したポイントクラウドに対し TDA を行った場合」について、ここで記載しているポイントクラウドの作成方法について説明する。幅 573 ピクセル、高さ 447 ピクセルの二値画像に対し、画像の左上のピクセルから順番に読み取っていく。このとき、黒ピクセル部分にあたる場所を読み取った際に、その場所の(幅, 高さ)を (x, y) として点を生成していくことで、 573×477 の点群を各画像においてそれぞれ作成している。

図 12 は以上のデータ作成の流れを簡単に表したものである。ここでは例として福岡県飯塚市の画像を挙げている。図中においてポイントクラウドの画像は、作成したポイントクラウドを視覚的にわかりやすくするためプロットした画像である。

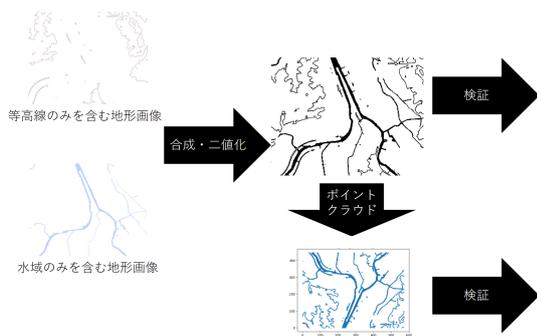


図 12: データ作成の流れ

また、各画像における水害レベルのラベル付けは、「ハザードマップポータルサイト」[11] を利用し行った。

5. 検証結果

以下に 4 つのアプローチにおける検証結果を順番に示す。

5.1 二値画像による位相的データ解析

まず本検証にて扱った画像について説明する。画像は、過去に水害が起きた地域、また解析に偏りが起きないように画像同士が似過ぎていない地域等を考慮して選択しており、計 30 枚用意した。また、画像に m0 から m29 と名前をつけている。m0 から m3 が福岡県朝倉市、m4 から m6 が福岡県飯塚市、m7 から m10 が福岡県久留米市、m11 から m15 が福岡県篠栗町、m16 から m20 が筑後川、m21 から m25 が福岡県直方市、m26 から m29 が山国川の画像となっている。また、節 4.1 にて説明したように、0 次元パーシステントホモロジー群についても調べたが、有用な情報を得ることができなかったため、1 次元パーシステント図

のみを用いることとした。

1 次元パーシステントダイアグラムをベクトル化する範囲は -5 から 35 としている。30 個の 1 次元パーシステント図に対してベクトル化によって得られた 30 個の 15 次元データに対して主成分分析を行った結果、第 1 主成分の寄与率が 99.87%、第 2 主成分の寄与率が 0.1%、第 3 主成分の寄与率が 0.014% であった。第 1 主成分の寄与率が 9 割以上と非常に高い値となっていることから、第 1 主成分にあたるパラメータの影響力が大きいことがわかった。ここで各主成分の重みについて確認すると、第 1 主成分についてはベクトル化した際の第 2 成分に当たる部分の重みが 100% であった(図 9 を参照)。また、第 2 主成分については第 5 成分の重みが 9 割以上の割合となっていた。

今回は寄与率が大きい第 1 主成分と対角線から最も離れた成分の重みが大きい第 2 主成分の二つの成分に注目して検証を行った。図 13 はこのときの主成分と主成分得点の散布図である。横軸が第 1 主成分、縦軸が第 2 主成分を表しており、赤い矢印(ベクトルの成分)は伸びている方向と矢印の長さで主成分得点を表している。

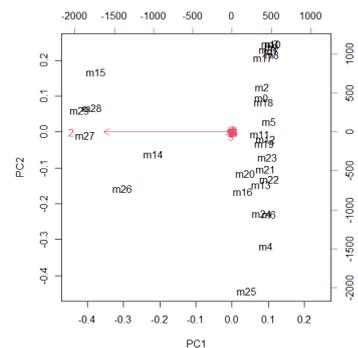


図 13: 主成分と主成分得点の散布図

主成分分析の結果に k-means 法を行った結果が図 14 である。クラスター数は 5 個の場合であり、横軸は第 1 主成分を、縦軸は第 2 主成分を表している。

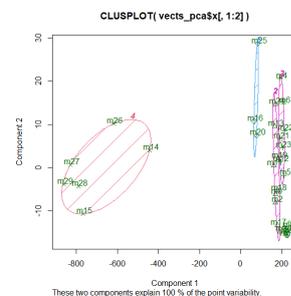


図 14: k-means 法の結果

実際に分類されたクラスター内での画像同士を目視にて比較してみると、画像の類似度で分類できており、主に等

高線の幅の類似度に基づくクラスターの分割が行われていた。つまり、第1主成分は等高線の幅であると考えられる。また、第2主成分は等高線の形などであると考えられるような結果であった。図15はその一部を抜粋したものである。つまり、第1主成分と第2主成分の内容から、画像の特徴による分類はできていることがわかった。従って、TDAによる地形画像からの特徴抽出は行えているといえる。しかし、クラスター分析の結果とハザードマップでの水害リスクを比較した場合、一部のクラスターでは全て同レベルのリスクであるものがみられたものの、ほとんどのクラスターには関係性を見出すことはできないような結果であった。

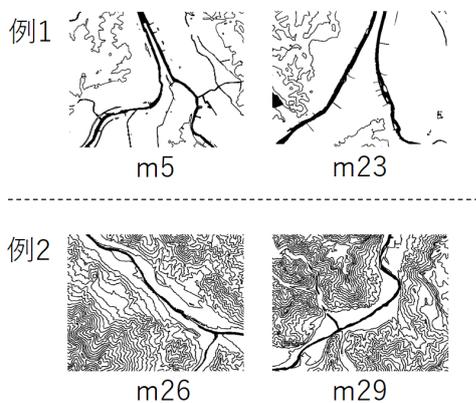


図 15: クラスター内の画像同士の例

5.2 ポイントクラウドによる位相的データ解析

ポイントクラウドは、節 4.2 にて説明したように、二値画像の黒ピクセル部分にあたる場所に点を生成することで作成している。画像は節 5.1 にて使用した 30 枚を利用しており、同様に 1 次元パーシステントダイアグラムのみを用いることとした。

また、変動させるベクトル化の birth, death の範囲は、0 から 100, 0 から 300, 0 から 500, 0 から 700, 0 から 900, 0 から 1100 の 6 通りで検証する。表 1 はそのときの結果をまとめたものである。第 1 主成分と第 2 主成分の寄与率の和が全て 9 割に近い値となっていたため、今回は第 1 主成分と第 2 主成分に注目していく。表 1 より、0 から 100, 0 から 300 では第 1 主成分の寄与率が 9 割を超えており、非常に高い割合となっている。また、0 から 500 以降では、第 1 主成分の寄与率が減少し、第 2 主成分の寄与率が増加していることがわかる。

表 1: 主成分分析の結果

ベクトル化の範囲	第 1 主成分の寄与率 (%)	第 2 主成分の寄与率 (%)
0 から 100	98.29	1.16
0 から 300	95.68	1.66
0 から 500	79.93	12.76
0 から 700	72.92	10.65
0 から 900	75.46	12.91
0 から 1100	77.65	9.19

各主成分の重みについて確認すると、第 1 主成分においては 6 通り全てにおいて、ベクトル化した際の第 2 成分に当たる部分の影響が大きいことがわかった (図 9 を参照)。つまり、対角線に近い要素が主に影響していることがわかる。第 2 主成分においては、主に第 4 成分, 第 5 成分, 第 9 成分の影響を受けていた。特に第 5 成分の重みが比較的大きいことから、対角線から離れた要素が主に影響していることがわかる。

図 16 は、左が二値画像に対しそのまま TDA 行った時の主成分分析の結果と、右が今回行った範囲 0 から 700 での主成分分析の結果の主成分と主成分得点の散布図である。横軸が第 1 主成分、縦軸が第 2 主成分を表しており、図中の赤い矢印 (ベクトルの成分) は伸びている方向と矢印の長さで主成分得点を表している。二値画像に対してそのまま TDA を行った場合では、第 1 主成分に対して矢印が一本しか見えない状況であったのに対し、ポイントクラウド化して TDA を行った場合では、複数の矢印が見られ、第 2 主成分に対しても矢印が確認できるようになっている。

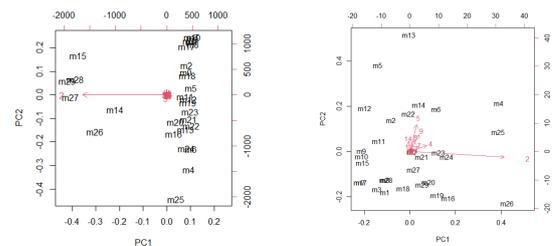


図 16: 主成分と主成分得点の散布図

以上の結果よりベクトル化の範囲は、第 1 主成分の寄与率が高すぎず、第 2 主成分の寄与率も考慮できる 0 から 500 以降の範囲が適していると考えられる。

5.3 ベクトル化に対して k -NN 法を用いた自然災害予測

節 5.2 にて得られた結果より、今回は二値画像の地形画像を基に生成したポイントクラウドにて TDA を行い、1 次元パーシステントダイアグラムのベクトル化範囲は 0 から 900 としている。また、0 次元パーシステントダイアグラムのベクトル化範囲は、無造作に選択した 20 個のデー

タに対し、最もデータ分布がまばらであった 0 から 600 にしている。

訓練データとして、水害リスクが大の画像データを 20 枚、中の画像データを 20 枚、小の画像データを 20 枚、計 60 枚のデータを用意した。つまり画像のラベルは、水害レベル大中小の 3 種類とする。テストデータは節 5.2 にて使用した 30 枚を用意している。表 2 は各 k の値における正答率である。

表 2: ベクトル化に対して k -NN 法を用いた結果

k の値	正答率 (%)
1	56.66
2	46.66
3	43.33
4	36.66
5	40
6	36.66
7	40
8	43.33
9	36.66
10	36.66

また、各 k のそれぞれのラベルにおける予測結果の正答率を表 3 に示す。

表 3: 各 k における予測結果の正答率

k の値	ラベル大の正答率 (%)	ラベル中の正答率 (%)	ラベル小の正答率 (%)
$k = 1$	70	45.45	55.55
$k = 2$	50	45.45	44.44
$k = 3$	50	27.27	55.55
$k = 4$	50	18.18	44.44
$k = 5$	60	18.18	44.44
$k = 6$	50	27.27	33.33
$k = 7$	50	27.27	44.44
$k = 8$	50	36.36	44.44
$k = 9$	50	27.27	33.33
$k = 10$	50	27.27	22.22

5.4 ボトルネックディスタンスでの k -NN 法を用いた自然災害予測

使用データは節 5.3 と同様のものを使う（訓練データ 60 枚、テストデータ 30 枚）。また、パーシステントダイアグラムに対し、そのままボトルネックディスタンスの k -NN 法を適用してしまうと、膨大な時間を要してしまうため、ベクトル化の時と同様ノイズ処理を行っている。今回はノイズ処理として重み付けではなく、death time - birth time ≤ 0.5 の要素、つまり生存時間が 0.5 以下の要素を全て削除するようにしている。表 4 は各 k の値における正答率である。

表 4: k -NN 法 (ボトルネックディスタンス) の結果

k の値	正答率 (%)
1	46.66
2	40
3	43.33
4	46.66
5	53.33
6	53.33
7	53.33
8	60
9	50
10	56.66

また、各 k のそれぞれのラベルにおける予測結果の正答率を表 5 に示す。

表 5: 各 k における予測結果の正答率

k の値	ラベル大の正答率 (%)	ラベル中の正答率 (%)	ラベル小の正答率 (%)
$k = 1$	50	27.27	66.66
$k = 2$	60	9.09	55.55
$k = 3$	40	36.36	55.55
$k = 4$	40	45.45	55.55
$k = 5$	60	45.45	55.55
$k = 6$	60	36.36	66.66
$k = 7$	60	36.36	66.66
$k = 8$	60	45.45	77.77
$k = 9$	60	45.45	44.44
$k = 10$	60	45.45	66.66

6. 結果・考察

二値画像に対してそのまま TDA を行った場合とポイントクラウドに対して行った場合両方において、主成分分析の結果より、地形情報が含まれる画像における等高線の幅や形などの特徴抽出が行われていることがわかった。特に図 16 から第 1 主成分と第 2 主成分に対してベクトルの各成分が強く表れていることから、二値画像に対しそのまま TDA を行うよりも、一度ポイントクラウドを作成して TDA を行う方が、特徴抽出を行いやすいと考えられる。

続いて自然災害予測等への応用ができるか検証することについてである。両方の k -NN 法の結果よりボトルネックディスタンスでの k -NN 法を用いた自然災害予測の方が正答率が高かった。また、ボトルネックディスタンスでの k -NN 法を用いた自然災害予測の手法では、 k の値が大きくなるほど正答率が高くなる傾向が見られた。

さらに、節 5.3 の表 3 と節 5.4 の表 5 より、水害リスクが大と小の画像に対しては正しい分類がされる傾向があるとわかる。これは特徴抽出での検証からもわかるように、

等高線の幅や形等が特徴的な場所は正しい分類がされやすいということを示しているといえる。

正答率が50%から60%程度であった理由として考えられる要因としては、検証に使用する地形画像データ数が少ないこと、画像のサイズや取得した等高線の幅が大きいもしくは小さい可能性があること、ベクトル化の次元数や重み付けなどのハイパーパラメータに問題があることなどが考えられる。

本研究では、使用した画像データ数が少ないことから自然災害予測への応用は期待される程度ではなかったが、地形情報を含む画像からの特徴抽出については行えていたことがわかった。画像データ数が少ないために十分な学習が行えず、50%から60%程度であったため、画像データ数を増やすことにより正答率が上がることを期待できる。

参考文献

- [1] 平岡裕章, タンパク質構造とトポロジー, 共立出版, 2013.
- [2] 林滯央, 位相的データ解析を用いたパターン認識の研究～手書き数字による実証～, 九州工業大学, 修士論文, 2020.
- [3] A. Zomorodian and G. Garsson, Computing Persistent Homology, *Discrete Comput. Geom.* Vol.33, 249-274(2005).
- [4] 大林一平, パーシステントホモロジーの計算ソフトウェアと発展的話題, 2017-02-28, 入手先 <https://i-obayashi.info/slides/lecture-ism-20170228.pdf>
- [5] 大林一平, 平岡裕章, Persistence Diagrams with Linear Machine Learning Models, *Journal of Applied and Computational Topology*, Volume1, page 421-449(2018), 05 May 2018, 入手先 <https://link.springer.com/article/10.1007/s41468-018-0013-5>.
- [6] HomCloud, 入手先 <https://homcloud.dev/index.html>
- [7] トポロジカルデータ解析コミュニティ, "Hom-Cloud(基礎編)", 2021-05-17, 入手先 [https://www.wpi-aimr.tohoku.ac.jp/TDA/members/files/Homcloud\(基礎編\).pdf](https://www.wpi-aimr.tohoku.ac.jp/TDA/members/files/Homcloud(基礎編).pdf)
- [8] 谷謙二, Web 等高線メーカー, Web 等高線メーカー—谷謙二研究室 (埼玉大学教育学部人文地理学), 入手先 <https://ktgis.net/service/webcontour/about.html>
- [9] 谷謙二 (2015) 標高タイルを利用した等高線作成 Web サイト「Web 等高線メーカー」の開発とそのアルゴリズム. 埼玉大学教育学部地理学研究報告, 35, 73-83.
- [10] 国土地理院, 地理院地図 Vector, 入手先 <https://maps.gsi.go.jp/vector/>
- [11] 国土地理院, ハザードマップポータルサイト, 入手先 <https://disaportal.gsi.go.jp/maps/>
- [12] 大森崇, 坂田真己子, 宿久洋, R Commander によるデータ解析 第2版, 共立出版, 2014.
- [13] 涌井良幸, 涌井貞美, 図解でわかる統計解析用語辞典, 初版, 日本実業出版社, 2003.
- [14] 金明哲, R によるデータサイエンス (第2版), 森北出版, 2017.