

地形画像の位相的特徴に基づく洪水リスク推定

下大田 憲世¹ 佐藤 好久²

概要: 近年、地形・地図などの大規模画像データから有用な構造情報を抽出し、防災へ活用する重要性が高まっている。本研究では、分布仮定に依存せずデータの形状の特徴を捉えられる位相的データ解析 (TDA) を用い、地形画像の位相的特徴に基づく洪水リスク推定が可能かどうかを検討した。等高線画像から点群を生成してパーシステンス図を算出し、1次元特徴のベクトル化範囲を選定したうえで kPCA により分離性を確認した。ベクトル化特徴による kNN 法とパーシステンスダイアグラムのボトルネックディスタンスによる kNN 法の2通りの方法で洪水リスク推定を行った。さらに今後はデータ拡充と設計の改善により推定性能の向上を図る。

キーワード: 情報数学、画像分類、機械学習

Flood Risk Estimation Based on Topological Features of Topographic Images

KENSEI SHIMOOTA¹ YOSHIHISA SATO²

Abstract: Recent disaster prevention requires extracting structural information from large-scale topographic and map images. This study investigates whether flood risk can be estimated from topological features of terrain images using topological data analysis (TDA), which captures shape without distributional assumptions. Contour images are converted to point clouds, persistence diagrams are computed, and the vectorization range of one-dimensional homological features is selected. Separability is examined using kernel PCA. Flood risk estimation is performed by kNN with vectorized features and by kNN with bottleneck distance between persistence diagrams. The results show feasibility in practice, and future work includes data expansion and improved design strategies.

Keywords: Information Mathematics-ALL, Machine learning, Image classification

1. はじめに

近年、情報技術の発展により、解析対象となるデータは大規模化・複雑化しており、人間が全体像を直感的に把握することが困難になっている。このような状況において、データに内在する特徴を効果的に抽出するための解析手法が重要視されている。近年、そのような手法の一つとして位相的データ解析 (TDA: Topological Data Analysis) が注目されている。TDA は、データの形状や連結性といった構造的特徴を位相的な観点から捉える解析手法であり、特定

の分布モデルを仮定せずに適用できる点に特徴がある。そのため、従来の統計的解析では扱いにくかった複雑な構造を持つデータに対しても有効であるとされている。地形情報を含む画像は、標高差や起伏の配置といった空間的構造を内包しており、これらの構造は洪水をはじめとする自然災害の発生と密接に関係していると考えられる。しかし、地形が持つ構造的特徴を適切に表現し、解析に利用することは容易ではない。そこで本研究では、地形情報を含む画像を対象として TDA を適用し、地形の構造的特徴を捉えることが可能であるかを検討する。また、得られた特徴を基に、水害を対象とした自然災害リスク推定への応用可能性について考察する。本研究を通じて、現地調査が困難な地域に対しても、地形画像に基づいた解析による災害予測

¹ 九州工業大学
Kyushu Institute of Technology
² 九州工業大学
Kyushu Institute of Technology

が可能になることを目指す。本研究は、増田圭汰による先行研究 [1] を継承しており、TDA を用いた洪水リスク予測手法のさらなる検証と改善を目的としている。

2. 位相的データ解析

本研究で用いる位相的データ解析の主要な概念であるパーシステントホモロジーについて述べる。

2.1 パーシステントホモロジーの概要

位相的データ解析では、データが持つ幾何学的構造を位相的不変量として捉える。特に重要な手法がパーシステントホモロジーであり、データの連結性や穴といった位相的特徴を定量的に記述することができる。

基本的なアイデアとして、データ点の集合に対してスケールパラメータを徐々に変化させながら、各スケールにおける位相的構造の変遷 (フィルトレーション) を追跡する。具体的には、各データ点を中心とする球体の半径を連続的に増加させていき、球体同士が接触した点で辺を、三つ以上の球体が共通部分を持った点で面を構成することで、複体と呼ばれる幾何学的構造を構築する。

このような構築過程において、連結成分の統合や閉曲線 (穴) の出現・消滅が生じる。それぞれの位相的特徴について、出現したスケール値を b (birth)、消滅したスケール値を d (death) として記録する。この (b, d) のペアの集合をパーシステントダイアグラムと呼ぶ。0次元の特徴は連結成分に、1次元の特徴は閉曲線構造に対応する。図1は、フィルトレーションの例を示している。

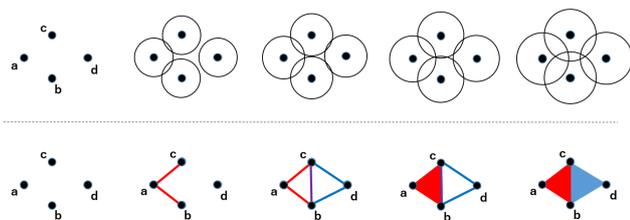


図1: フィルトレーションの例

パーシステンスダイアグラムは、横軸に b 、縦軸に d をとった平面上の点集合として視覚化される。この表現をパーシステンス図と呼び、対角線 $y = x$ から離れた点ほど長期間継続した特徴であり、データの本質的な構造を反映していると考えられる。一方、対角線近傍の点は短命な特徴であり、ノイズと見なされることが多い。

3. 機械学習

3.1 主成分分析

変数が少ないデータの場合、そのデータの構造を明らかにすることは難しくない。しかしながら、変数が大きくなればなる程、構造が複雑になり、解析が難しくなる。一方、

変数が多くなると、相関関係がある変数が存在する可能性もあり、そのデータの構造に対して「本質的な」変数を取り出す重要性が増す。

主成分分析 (PCA: Principal Component Analysis) は、多変量データを少数の総合的な指標に集約する手法である。複数の変数間に存在する相関を活用し、データに含まれる情報を可能な限り保持したまま、互いに無相関な新しい変数 (主成分) を構成することで、データ構造の理解を容易にする (図2)。

主成分分析における**主成分**は、元の変数を線形結合して得られる合成変数である。第1主成分は、全体の分散が最大となるように各変数の重みを決定して構成される。分散が大きい変数ほどデータの特徴を捉えやすいという考え方に基づいている。続く第2主成分は、第1主成分と統計的に独立 (直交) となり、かつ残された分散が最大になるように構成される。第3主成分以降も同様に、既存の主成分と直交し分散が最大になる条件のもとで順次作成される。この手続きにより、元の変数と同じ数だけ主成分を得ることが可能である。

図2は、主成分分析による次元削減の例である。図2の左図は2変数で表現されたデータの分布を示している。右図では、このデータに対して分散最大となる方向に第1主成分を、それと直交する方向に第2主成分を設定している。第1主成分のみでデータの主要な変動が捉えられていることから、2次元データを1次元に縮約しても十分な情報が保持されることがわかる。

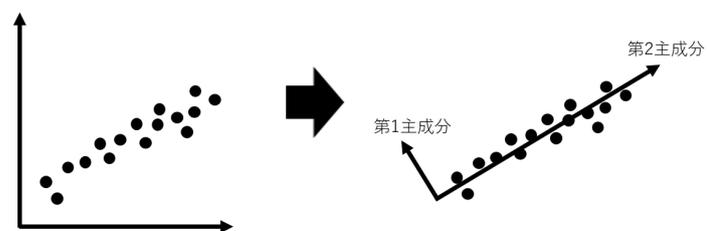


図2: 主成分分析による次元削減の例

3.2 カーネル主成分分析

通常的主成分分析に対して、**カーネル主成分分析 (kPCA: Kernel Principal Component Analysis)** は、非線形的なデータ構造を持つ場合に有効である。線形的なデータ構造を仮定する通常的主成分分析では非線形的なデータの解析の手法は使えない。非線形的なデータ構造を何らかの方法で線形的なデータ構造に変換することができれば、線形的なデータ解析の手法を適用することができる。このような考え方により、「カーネル関数」を用いてデータを表現し直し線形的な手法を行うことを**カーネル法**という。**カーネル主**

成分分析のアルゴリズムの主な流れは以下の通りである：

- (1) カーネル関数 $K(x, y)$ を決める
- (2) カーネル関数 $K(x, y)$ により表現し直されたデータを基に、[共分散行列] 等関連する行列 $K_{n \times n}$ を定める
- (3) 正方行列 $K_{n \times n}$ の固有値、固有ベクトルを求める
- (4) 固有値、固有ベクトルから得られる情報を抽出する

3.3 クラスタ分析

クラスタ分析とは、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。個体間の類似度あるいは非類似度(距離関数)に基づいて最も似ている個体から順にまとめてクラスタを構成する「階層的クラスタ分析」と、そうではない「非階層的クラスタ分析」がある。階層的クラスタ分析では目的に応じて距離関数を選択できるが、個体数が多いと計算量が膨大になり大規模データには不向きである。一方、大量のデータ分析には非階層的クラスタ分析が多く用いられ、代表的なものとして k-means 法がある。

図3は、k-means 法の手順を簡単に表したものである。k-means 法は、いくつかクラスタ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の中心との距離を求め、データを最も近いクラスタに振り分け直し、新しく作成したクラスタの中心を求める。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。

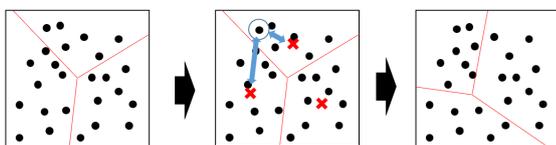


図3: 非階層クラスタ分析の手順

3.4 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う。非線形判別分析では、1次関数以外の判別関数を用いる方法、距離に基づいた判別方法、多数決による判別方法、ベイズ判別法など多くの方法がある。

このうち、多数決による判別方法の一つとして k 最近傍 (k -Nearest Neighbor) 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。

図4は $k = 5$ の場合である。未知データである \star を中心に最も近いものを5個探している。図中円内に \bullet が1個、 \blacktriangle が3個、 \blacksquare が1個であることから、未知データ \star は \blacktriangle のグループに属すると判断される。

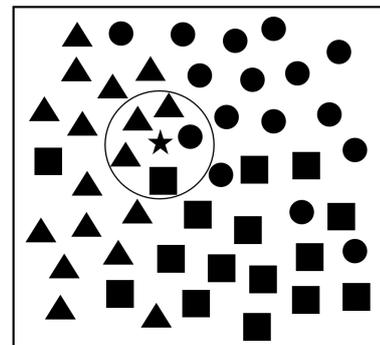


図4: k-NN 法の例

4. 実験

4.1 目的

本研究では、地形情報を含む画像から TDA で特徴抽出が可能かを検証し、その知見を洪水リスクの推定に活用できるかを確認する。多様な災害を一度に扱うのは現実的でないため、対象を洪水に絞って検討を進める。

最終的には、地上調査が難しい地域や他惑星などにおいても、画像データのみから災害リスクを推定できる手法の足がかりを得ることを目指す。

4.2 検証方法

本検証では以下の流れで解析を実施する。

- (1) 本研究で使用するデータを作成する。ここで洪水リスクに応じたラベル付けを実施。
- (2) 作製したデータに対して TDA を適用し、0次元と1次元のパーシステンスダイアグラムを取得する。
- (3) 得られたパーシステンスダイアグラムに対して、PCA を用いて複数のベクトル化範囲を試行し、最適な範囲を決定する。
- (4) 決定された範囲でベクトル化した特徴量に対して kPCA を実施する。
- (5) kPCA の結果をクラスタ分析で可視化し、地形特徴の抽出可能性を検証する。
- (6) k -NN 法により予測を実施する。

- (i) ベクトル化した特徴量を用いた予測 (手法 1) を実施する。
- (ii) パーシステンスダイアグラム間のボトルネックディスタンスを利用した予測 (手法 2) を実施する。

4.3 データ作成

本研究で使用するデータは、「等高線メーカー」サイト [2] と地理院地図 Vector[4] を基に作成している。まず等高線メーカーより等高線の情報を取得し、地理院地図 Vector より水域 (河川や池) の情報を抽出する。このとき、水域を等高線よりも強調している。次に、図 5 に示すように、取得した等高線と水域の情報を合成し、二値化処理により地形画像を生成する。二値化画像では、黒ピクセルが等高線と水域、それ以外が白ピクセルに対応している。

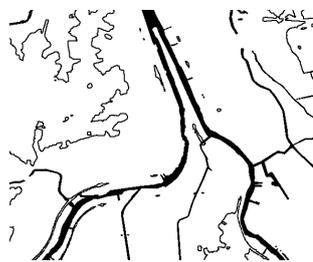


図 5: 二値化画像の例

最後に、図 6 に示すようなポイントクラウドを生成するため、二値化画像から黒ピクセルの座標 (x, y) を抽出し、各ピクセル位置をポイントとして記録する。黒ピクセルの抽出は、画像の左上 $(0, 0)$ から始まり、 y 方向 (行方向) に順次走査し、その後 x 方向 (列方向) に進む列優先 (column-major) 順で行われる。すなわち、 $(0, 0), (0, 1), (0, 2), \dots, (0, H-1), (1, 0), (1, 1), \dots$ という順序で座標が記録される。このポイントクラウドが TDA の入力となり、地形の幾何学的特性を反映した永続ホモロジーの計算が行われる。

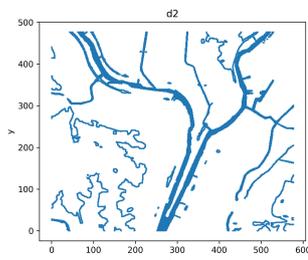


図 6: ポイントクラウドの例

4.3.1 ラベル付け

各地形画像に対する洪水リスクレベルのラベル付けは、「ハザードマップポータルサイト」 [5] から取得したハザ-

ードマップ情報を利用する。ハザードマップは浸水深を色で表現しており、黄色系 (浸水深 1~3m)、オレンジ系 (浸水深 3~5m)、赤系 (浸水深 5m 以上) などの色分けがなされている。取得した各ハザードマップ画像に対して、各ピクセルをスコア (黄色=1 点、オレンジ=2 点、赤=3 点、その他=0 点) に変換する。この総スコアを計算し、総ピクセル数で割って平均スコアを計算する。

この平均スコア値に基づいて、以下の 3 つのカテゴリにラベル付けする：

- (1) 平均スコア ≥ 1.0 の場合：「大」 (大規模浸水)
- (2) $0.3 \leq$ 平均スコア < 1.0 の場合：「小」 (小規模浸水)
- (3) 平均スコア < 0.3 の場合：「無」 (浸水なし)

先行研究の「大・中・小」3カテゴリから「大・小・無」の3カテゴリに変更したのは、予測対象を実務的な意思決定に直結する「浸水がある/ない」「浸水規模が大きい/小さい」という判別に焦点を当てるためである。

4.4 パーシステンスダイアグラムのベクトル化

得られた 0 次および 1 次のパーシステンスダイアグラム (PD0, PD1) を機械学習の入力とするため、固定長のベクトルに変換する。本研究では、birth 値と death 値の範囲を PCA により決定し、その範囲を 5×5 のグリッドに分割する (図 7)。

1 次ホモロジー (H_1) については、birth 軸と death 軸をそれぞれ 5 分割し、上三角領域 ($\text{death} \geq \text{birth}$) の 15 個のセルに対応する 15 次元ベクトルを生成する。0 次ホモロジー (H_0) については、death 軸のみを 5 分割した 5 次元ベクトルを生成する。これらを結合し、合計 20 次元の特徴ベクトルとする (図 8)。

各点 (b, d) に対して、persistence $(l = d - b)$ を計算し、重み $w = \arctan(0.01 \cdot l^3)$ を付与する。この重み付けにより、長寿命な特徴 (persistence が大きい点) がより強調される。各点が属するセルに重みを加算することで、最終的な特徴ベクトルが得られる。

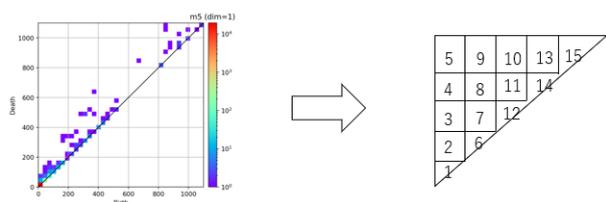


図 7: 1 次のパーシステンスダイアグラムのベクトル化の例

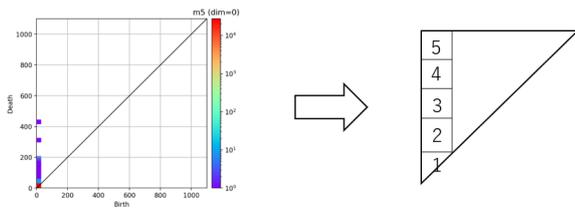


図 8: 0 次のパーシステンスダイアグラムのベクトル化の例

5. 結果

5.1 位相的データ解析

kPCA を適用する前に、PCA を用いてベクトル化範囲を決定した。

まず、 H_0 のベクトル化範囲を決定するため、 $\max v$ (最大値) を 100, 300, 500, 700, 900 の 5 パターンで試行し、それぞれについて PCA を実施した。表 1 に各範囲における主成分の寄与率を示す。

表 1: PCA 寄与率分析 ($H_0:5$)

ベクトル化範囲	PC1 寄与率	PC2 寄与率	累積 (PC1+PC2)
0~100	0.380	0.238	0.618
0~300	0.459	0.217	0.677
0~500	0.458	0.203	0.661
0~700	0.406	0.235	0.641
0~900	0.383	0.218	0.601

H_0 では、連結成分が統合されていく過程から「局所的な密度による結合」と「塊同士が結合する距離スケール」の両方を捉える必要がある。そのため、両者を過不足なく含みつつ次元圧縮時の情報保持も高水準 (Cum2=0.661) である $\max v=500$ を採用した。

次に、 H_1 のベクトル化範囲を決定するため、 $\max v$ (最大値) を 100, 300, 500, 700, 900 の 5 パターンで試行し、PCA を実施した。このとき、決定した範囲でベクトル化した H_0 のベクトルを加えた 20 次元のベクトルで実施している。表 2 に各範囲における主成分の寄与率を示す。

表 2: PCA 寄与率分析 ($H_1:15 + H_0:5$)

ベクトル化範囲	PC1 寄与率	PC2 寄与率	累積 (PC1+PC2)
0~100	0.615	0.193	0.808
0~300	0.422	0.158	0.580
0~500	0.293	0.245	0.538
0~700	0.276	0.219	0.495
0~900	0.255	0.183	0.438

表 2 から、以下のことが読み取れる。0~100 の範囲では、PC1 が全体の 6 割以上を占める支配的な成分となっており、情報が潰れ気味で、ほぼ 1 次元的な構造しか捉えら

れていない。これに対して 0~300 では、累積寄与率がまだ高い (58%) もの、PC1 への偏重が緩和される安定・保守的な範囲となっている。

最も注目すべき点は、0~500 の範囲である。この範囲では、PC1 と PC2 がほぼ拮抗 (それぞれ 29.3% と 24.5%) しており、多次元構造が最もきれいに表現される。一方、0~700 および 0~900 では、累積寄与率が 5 割未満に低下し、情報が高次元に分散しすぎてノイズの影響が増す傾向が見られる。

これらの分析結果から、PCA によるベクトル化の最適な範囲は、0~500 である。この範囲は、地形の「集約性」と「表現力」のバランスが最も良く取られており、洪水予測タスクに適した特徴表現を実現すると考える。

次に、パーシステンスダイアグラムをベクトル化した特徴量に対して、より複雑な非線形構造を捉えることを目的として、kPCA を適用する。

図 9 に、kPCA により 2 次元に圧縮された特徴空間を示す。この空間にクラスター分析 (k -means、クラスター数 $k=3$) を適用することで、地形特徴の違いを明確に可視化した。kPCA により、地形の複雑な非線形パターンが効果的に分離され、同一の地形特徴を持つ地形グループが抽出されている。

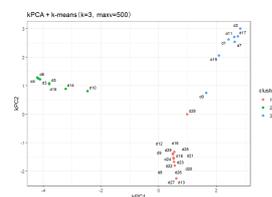


図 9: kPCA と k -means クラスタリングによる地形特徴の可視化 ($k=3$)

表 3 に、kPCA により射影された各データの割り当てられたクラスター番号を示す。

表 3: k -means クラスタリングの結果

sample	cluster	sample	cluster	sample	cluster
d0	3	d10	2	d20	1
d1	3	d11	3	d21	1
d2	3	d12	1	d22	1
d3	2	d13	1	d23	1
d4	2	d14	2	d24	1
d5	2	d15	2	d25	1
d6	2	d16	1	d26	1
d7	3	d17	3	d27	1
d8	1	d18	1	d28	1
d9	1	d19	3	d29	1

5.2 ベクトル化に対して k -NN 法を用いた洪水予測

パーシステンスダイアグラムをベクトル化した 20 次元特徴量 (H_0 :5 次元 + H_1 :15 次元) に対して、 k -NN 法を適用した。1 次元パーシステンスダイアグラムのベクトル化の範囲は 0~500、0~1000、0~1500 の 3 種類を試行し、最も高い正答率を示した 0~500 を採用した。また、0 次元パーシステンスダイアグラムのベクトル化の範囲は、データ分布が最もまばらである 0~500 を採用した。訓練データは 180 個 (「大」「中」「小」各 60 個) を使用し、テストデータは 30 個 (各クラス約 10 個) を使用している。 $k = 1 \sim 10$ で分類精度を比較した結果 (表 4) と各 k の混同行列 (表 5~表 14) を以下に示す。

表 4: ベクトル化特徴量による予測の結果 (正答率)

k	結果 (正答率)
1	0.533
2	0.633
3	0.533
4	0.700
5	0.700
6	0.733
7	0.633
8	0.600
9	0.633
10	0.600

表 5: $k = 1$ における予測結果

	大	小	無
大	5	3	3
小	3	5	1
無	1	3	6
正答率: 0.533			

表 6: $k = 2$ における予測結果

	大	小	無
大	7	2	2
小	3	4	2
無	2	0	8
正答率: 0.633			

表 7: $k = 3$ における予測結果

	大	小	無
大	5	4	2
小	2	5	2
無	1	3	6
正答率: 0.533			

表 8: $k = 4$ における予測結果

	大	小	無
大	9	1	1
小	3	5	1
無	2	1	7
正答率: 0.700			

表 9: $k = 5$ における予測結果

	大	小	無
大	9	1	1
小	3	5	1
無	1	2	7
正答率: 0.700			

表 10: $k = 6$ における予測結果

	大	小	無
大	9	1	1
小	3	5	1
無	1	1	8
正答率: 0.733			

表 11: $k = 7$ における予測結果

	大	小	無
大	9	1	1
小	3	5	1
無	0	5	5
正答率: 0.633			

表 12: $k = 8$ における予測結果

	大	小	無
大	9	1	1
小	5	3	1
無	0	4	6
正答率: 0.600			

表 13: $k = 9$ における予測結果

	大	小	無
大	9	1	1
小	3	5	1
無	0	5	5
正答率: 0.633			

表 14: $k = 10$ における予測結果

	大	小	無
大	9	1	1
小	4	4	1
無	0	5	5
正答率: 0.600			

5.3 ボトルネックディスタンスでの k -NN 法を用いた洪水予測

パーシステンスダイアグラム間のボトルネックディスタンスに対して k -NN 法を適用した。ここでも訓練データとテストデータは、節 5.2 と同様に 180 個（「大」「小」「無」各 60 個）と 30 個（各クラス約 10 個）を使用している。ここで、パーシステンスダイアグラムのノイズ除去のため、生存時間 (death-birth) が 0.5 以下の点は除去した。

$k = 1 \sim 10$ で分類精度を比較した結果（表 15）と各 k の混同行列（表 16～表 25）を以下に示す。

表 15: ボトルネックディスタンスによる予測の結果（正答率）

k	結果（正答率）
1	0.400
2	0.467
3	0.400
4	0.533
5	0.567
6	0.467
7	0.433
8	0.467
9	0.467
10	0.500

表 16: $k = 1$ における予測結果

	大	小	無
大	3	4	4
小	5	2	2
無	0	3	7
正答率: 0.400			

表 17: $k = 2$ における予測結果

	大	小	無
大	6	2	3
小	6	0	3
無	1	1	8
正答率: 0.467			

表 18: $k = 3$ における予測結果

	大	小	無
大	5	3	3
小	5	1	3
無	2	2	6
正答率: 0.400			

表 19: $k = 4$ における予測結果

	大	小	無
大	4	4	3
小	3	2	4
無	0	0	10
正答率: 0.533			

表 20: $k = 5$ における予測結果

	大	小	無
大	4	3	4
小	2	3	4
無	0	0	10
正答率: 0.567			

表 21: $k = 6$ における予測結果

	大	小	無
大	3	4	4
小	3	2	4
無	1	0	9
正答率: 0.467			

表 22: $k = 7$ における予測結果

	大	小	無
大	3	4	4
小	4	2	3
無	2	0	8
正答率: 0.433			

表 23: $k = 8$ における予測結果

	大	小	無
大	3	4	4
小	3	3	3
無	2	0	8
正答率: 0.467			

表 24: $k = 9$ における予測結果

	大	小	無
大	4	4	3
小	2	3	4
無	1	2	7
正答率: 0.467			

表 25: $k = 10$ における予測結果

	大	小	無
大	4	3	4
小	2	3	4
無	1	1	8
正答率: 0.500			

6. 結果と考察

6.1 結果の要約

ベクトル化特徴量に対する k -NN (手法 1) では $k = 6$ で正答率 73.33%と最良の結果が得られた。一方、ボトルネックディスタンスに基づく k -NN (手法 2) は $k = 5$ で 56.67%に留まり、ベクトル化特徴量の方が高い識別性能を示した。精度の向上は見られなかったが、手法 2 では「無」クラスの再現率が高く、 $k = 4$ と $k = 5$ では、100%であった。

6.2 考察

本研究では、kPCA により 2 次元に圧縮された特徴空間において、等高線の幅や形態が類似する地形画像を分類可能であることを確認した。線形 PCA では分離が明確でなかったが、非線形写像により、サンプル間の差異がより明瞭に表現された。

この結果から、TDA により、地形の複雑な非線形パターンを表現できることが示された。特に、kPCA による非線形写像により、地形の幾何的差異を明確に表現でき、地形の複雑な構造を保持したまま低次元空間への射影が実現できることが示された。

k -NN 法による予測では、ベクトル化特徴量を用いた手法 1 が $k = 6$ で最高正答率 73.33%を達成した一方、ボトルネックディスタンスを用いた手法 2 では $k = 5$ で 56.67%に留まった。手法 1 は手法 2 と比較して約 16%ほど精度が高く、ベクトル化による特徴表現の有効性が確認された。また、手法 2 は「無」クラスの判別に偏る傾向が見られ、全体的なバランスの取れた分類性能には課題が残るが、「無」クラスに対する再現率は 100%であり、「無」の検出能力は非常に高いことが示された。

先行研究 [1] と比較すると、ボトルネックディスタンスを用いた手法 2 では精度向上は見られなかったが、手法 1 は約 10%の精度向上が確認できた。これは、訓練データ数を 3 倍ほど増やしたことに加え、ラベル付けを「大・中・小」から「大・小・無」に変更して判別基準を実務的な意思決定に合わせたことなどが原因と考えられる。

手法 1 が手法 2 を大幅に上回った理由として、以下の 3 点が挙げられる。第一に、ベクトル化では、パーシステンスダイアグラムを 20 次元の数値ベクトルに変換する。各次元はグリッドの各セルに含まれる点の個数を表すため、

地形特徴の空間的分布が詳細に保持される。これに対してボトルネックディスタンスは 2 つのパーシステンスダイアグラムの類似度を 1 つの数値で表現するため、点の分布に関する詳細な情報が失われやすい。第二に、ボトルネックディスタンスは図全体の大域的な類似度を測定するため、洪水リスク判別に重要な局所的な地形特徴の違い（特定の高さ範囲での連結成分の変化など）を十分に捉えられなかったこと。第三に、複数のベクトル化の範囲 (0~500) を試し、選択することにより、データ分布の特性に適した特徴抽出ができたことである。

本研究の課題として、「小」クラスの地形特徴が「大」「無」と重なりやすく、境界が曖昧で「小」クラスの誤分類率が高く、クラス間の判別に改善の余地があること、ボトルネックディスタンスによる手法の性能向上の余地が大きく距離指標の工夫が必要であることが挙げられる。今後の改善方策として、訓練・テストデータ規模の拡大、異なる距離指標の検討、さらにはディープラーニングなどの高度な機械学習手法の導入が考えられる。これらにより、洪水予測精度のさらなる向上が期待される。

7. おわりに

本研究は、地形画像から得られる位相的特徴を洪水リスク推定に結び付ける枠組みを示した。今後は、地域や時期の異なるデータへの適用やマルチモーダル技術の導入、特徴設計と評価指標の再検討を進め、実運用に耐える判断材料としての有効性を高める。

参考文献

- [1] 増田圭汰, 「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」, 九州工業大学大学院修士学位論文, 2023 年.
- [2] 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), <https://ktgis.net/service/webcontour/about.html>
- [3] 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.
- [4] 国土地理院, 地理院地図 Vector, <https://maps.gsi.go.jp/vector/>
- [5] 国土地理院, ハザードマップポータルサイト, <https://disaportal.gsi.go.jp/maps/>