

自然の多面的価値の理解に向けたランドスケープの埋め込み表現の開発

芳賀 智宏

大阪大学 大学院工学研究科

1. はじめに

景観生態学をはじめとする環境分野の地理空間情報解析では、空間に存在する地物の種類と配置を分析し、人間社会と自然生態系を包含する社会・生態システムの構造と機能を明らかにしてきた。近年では、自然の多面的価値を評価することの重要性が認知され、特定の生物種の分布の予測するような自然の内在的価値中心の研究に加えて、自然が人間社会にもたらす道具的価値や、人間と自然の関わりの中で生まれる関係価値など、質的側面の評価が求められている [1]。

地理空間情報解析で一般に用いられる土地利用・被覆のデータは、メッシュやポリゴン単位でそこに存在する地物の情報を要約して一つのクラスラベルを付与する抽象化されたデータである。このため、自然環境調査の成果物である環境省の植生図では人工的土地利用の分類が少ない、国土利用の調査で用いられる国土数値情報の土地利用細分メッシュでは自然生態系の情報が少ないなど、データセットの作られた目的によってどの項目をどの程度抽象化するかが異なる。このため、社会システムと自然生態システムの横断的に解析するには、一つの情報源のみで双方のシステムの情報を網羅することが困難である。

近年は OpenStreetMap (OSM) など、個別の地物の情報をポイント・ライン・ポリゴン情報として保持した大規模かつオープンなデータが公開されるようになってきた。しかし、カテゴリ数の多さや、集計した際に疎ベクトルになるなどの技術的な課題から、環境分野での利用は限定的である。

このギャップを埋めるため、Hex2Vec [2] や GeoVeX [3] など、表現学習を応用し、大規模かつ詳細な地理空間情報を低次元の密ベクトル化する手法が提案され、下流タスクでの優位性を示し

た [2, 3]。また、地物名をグリッドの中心からの距離に応じて並べることで、そのグリッドについての疑似的な文章を作成し、BERT で埋め込みベクトルを作ることも提案されている [4]。

本研究では、社会・生態システムのランドスケープの地理空間情報を統合した埋め込みベクトルを作成し、1. 気候変数の推定、2. 人口・地価の推定、3. 継続的な市民科学活動の実施の有無、4. 保全地域で重視されている自然の価値の分類の4つの下流タスクで性能を評価した。

2. 方法

2.1 使用したデータの説明

地域の社会・生態システムの空間的な状態を統合的に評価するために、自然植生と人工的地物の情報を両方用いた。自然植生の情報については、日本全国標準土地利用メッシュデータ (LULC) を用いた [5, 6]。このデータは、環境省の5万分の1植生図をもとに、自然植生に関わる土地利用・被覆は45種類、人工的土地利用は4種類のカテゴリに分類した1 km 解像度のメッシュデータである。また、人工的な地物の情報が豊富に記録されているOSMの12,404種類のタグから先行研究を参考に316タグを抽出した。これらの2つの種類のデータを、Uber H3 grid system の resolution 8 (約1×1 km 解像度) で集計した。具体的には、LULCではグリッド内の49種類の土地利用・被覆カテゴリ別の面積割合、OSMではグリッド内の316タグ別の出現個数を集計し、日本全国597,109グリッドに対して、355次元のデータセットを作成した。

2.2 モデルの説明

OpenStreetMap から埋め込みベクトルを作成することができるHex2VecとGeoVeXを用いた。

Hex2Vec は word2vec を参考に開発されたモデルである。Skip-gram with negative sampling を用いて、任意の 2 つのグリッドについて空間的に隣接しているかを 2 値分類するタスクを学習する過程で、各グリッドの埋め込みベクトルを獲得する [2]。一方で GeoVeX は、畳み込みニューラルネットワーク (CNN) を用いたオートエンコーダであり、注目しているグリッドに加えて周辺のグリッドの情報を参照して 32 次元の埋め込みベクトルを獲得する [3]。

本研究では、日本全国の 597,109 グリッドを対象とした。A. LULC のみを用いた場合 (LULC)、B. OSM のみを用いた場合 (OSM)、C. LULC と OSM を列方向に結合した場合 (統合) の 3 通りの入力データセットを作成し、Hex2Vec と GeoVeX で埋め込みベクトルを作成した。各モデルの epoch 数は 128、バッチサイズは Hex2Vec は 2048、GeoVeX は 512、学習率は 0.001 とした。埋め込みベクトルの次元数にはデフォルトを採用し、Hex2Vec は 50、GeoVeX は 32 次元とした。

2.3 埋め込みベクトルの特性の評価

3 種類の入力データセットと 2 種類の埋め込みモデルで作成した 6 種類の埋め込みベクトルについて、空間分布と埋め込みベクトル軸の解釈可能性を評価した。先行研究を参考に、各埋め込みベクトルを独立成分分析で 3 次元に削減したのちに、各軸を 0–255 にスケールして RGB 画像として可視化し、空間的な特性を視覚的に評価した。その際、自然言語分野の単語埋め込みベクトルの解釈で用いられる手法を参考に、独立成分分析後の 3 次元の各軸について上位 100 個のグリッドを抽出し、出現頻度の高い特徴量を確認した [7,8]。これにより、人工的土地利用を表す軸には R、二次的土地利用を表す軸には G、自然性の高い土地利用を表す軸には B を割り当てた。

2.4 下流タスクを用いた応用可能性の評価

地理空間情報の埋め込みベクトルの先行研究でよく用いられる下流タスクである 1. 年平均気温・年降水量の回帰、2. 人口・地価の回帰に加えて、生態系保全に関連する下流タスクとして、3.

継続的な市民科学活動の実施の有無の分類、4. 環境省が認証した自然共生サイトで認証された自然の価値・機能の分類の合計 4 種類の下流タスクに対して、埋め込みベクトルの性能を評価した。A. LULC データセットを説明変数として用いた結果をベースラインとして、入力データセットと埋め込みモデルの組合せごとに性能を評価した。分類タスクではロジスティック回帰モデル、回帰タスクでは、重回帰モデルを用いた。日本の地方 9 区分で空間ブロッキングを施した交差検証を行い、交差検証誤差でモデルの性能を評価した。

3. 結果

3.1 埋め込みベクトルの特性の評価

図 1 に各埋め込みベクトルを可視化した結果を示す。3 種類の入力データセットを FastICA のみで 3 次元に削減した結果からは、A. LULC データセットでは都市部と森林の違いが明確に現れた。A. LULC データセットでは 49 種類のラベルのうち 45 種類が植生の分類であるため、都市部よりも森林植生の分類の空間的特徴が強調された。一方で、B. OSM および C. 統合データセットを FastICA で次元削減した場合、各軸の上位 100 個の地物間の特徴の違いを定性的に解釈することができず、図 1 から、次元削減の過程で情報が損失されたことが示された。

Hex2Vec と GeoVeX モデルによる埋め込みベクトルを用いた場合、A. LULC データセットを用いた場合は森林植生が、B. OSM データセットを用いた場合は都市部が詳細に分類された。C. 統合データセットを用いた場合は、自然生態系と都市の両方が詳細に分類された。特に Hex2Vec では、A. C. のデータセットでは、西日本の暖温帯常緑広葉樹林、東日本から東北にかけての冷温帯落葉広葉樹林、北海道の針広混交林へと連続的に変化するという日本の森林植生の特徴を捉えた埋め込みベクトルが作成された。GeoVeX による埋め込みでも、A. C. のデータセットで同様の森林植生の違いが表現されたが、Hex2Vec よりも地域間の違いは小さかった。

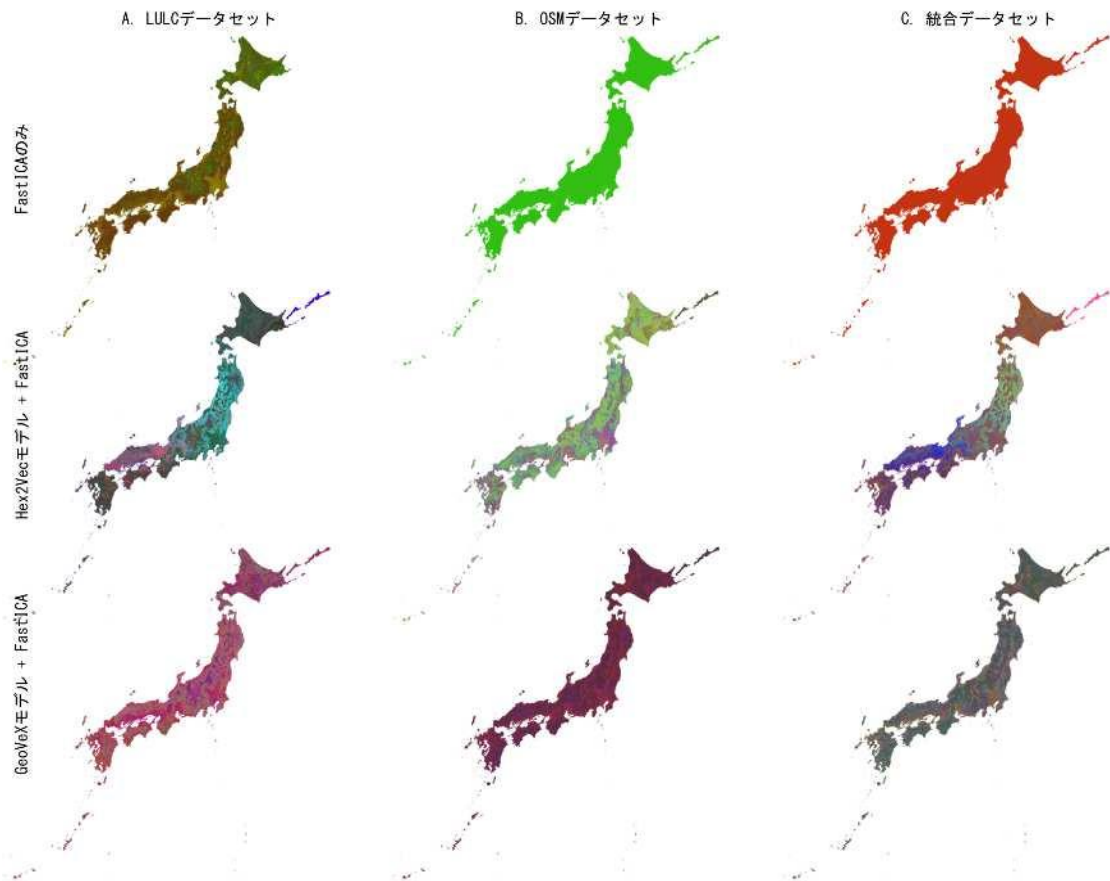


図 1. 各埋め込みベクトルの視覚的比較

3.2 下流タスクでの性能評価

下流タスクの交差検証誤差を表 1 に示す。全ての下流タスクで、埋め込みベクトルを用いた場合に性能が向上したが、最も性能が高かった埋め込みモデルと入力データセットの組み合わせはタスクによって異なった。気候変数の回帰タスクでは、Hex2Vec- 統合データセット、社会・経済変数の回帰タスクと、定期的な生物調査の有無の分類タスクでは、Hex2Vec-OSM データセットの組合せが最も高い性能を示した。また、環境省の自然共生サイトの認証の有無や、認証の際に自然の価値のどの側面を重視したかを分類するタスクでは、GeoVeX を用いた埋め込みベクトル群が高い性能を示した。

4. 考察

4.1 埋め込みベクトルの特性

3.1 の結果から、OSM や日本全国標準土地利用メッシュのような、分類が詳細な地理空間情報

データに対しては、一般的な次元削減の手法よりも、表現学習を用いた埋め込みベクトルを作成することで各グリッドの特徴を情報損失を抑えて低次元に削減することができる事が示された。

また、埋め込みベクトルの作成に用いるモデル内部のアルゴリズムの違いによって、埋め込みベクトルの空間的性質が変化することも明らかになった。Hex2Vec では、*"Everything is related to everything else, but near things are more related than distant things"* という *"first law of geography"* に従って、各グリッドが隣接しているかどうかを学習している [2, 9]。一方で GeoVeX では、入力データとデコーダの出力の誤差を評価する際に、中心のグリッドからの距離に応じて誤差に重みをかける方法を採用している。どちらも地理空間的データの空間分布に配慮しているが、入力データの大域的特徴を学習する CNN を用いたオートエンコーダの構造を採用した GeoVeX では、隣接するグリッドの埋め込みベクトルの特徴量が類似し、

表 1. 下流タスクでの交差検証誤差

モデル・データ	Top1の タスク	回帰タスク (RMSE)					分類タスク (AUC)				
		気候		社会・経済			生物調査		環境省の認証制度		
		年平均気温	log10(年降水量)	log10(人口)	log10(地価+1)		定期的な調査	認証の有無	理由1. 重要な場	理由2. 重要な種	理由3. 重要な機能
ICA - 土地利用	0	3.36 ±1.33	0.15 ±0.05	0.81 ±0.27	0.57 ±0.29		0.61 ±0.08	0.63 ±0.14	0.63 ±0.16	0.62 ±0.19	0.58 ±0.22
ICA - OSM	1	4.22 ±1.6	0.15 ±0.05	0.76 ±0.08	0.52 ±0.13		0.68 ±0.07	0.69 ±0.1	0.69 ±0.12	0.71 ±0.14	0.52 ±0.32
ICA - 統合	0	4.30 ±1.62	0.14 ±0.04	0.68 ±0.17	0.46 ±0.13		0.63 ±0.07	0.68 ±0.11	0.70 ±0.15	0.67 ±0.15	0.46 ±0.36
hex2vec - 土地利用	0	2.20 ±0.94	0.13 ±0.04	0.52 ±0.06	0.41 ±0.09		0.75 ±0.07	0.56 ±0.17	0.60 ±0.18	0.65 ±0.21	0.73 ±0.22
hex2vec - OSM	3	3.50 ±1.34	0.15 ±0.05	0.49 ±0.04	0.36 ±0.08		0.80 ±0.07	0.63 ±0.12	0.61 ±0.07	0.57 ±0.12	0.53 ±0.26
hex2vec - 統合	2	2.15 ±0.94	0.12 ±0.05	0.52 ±0.09	0.41 ±0.15		0.76 ±0.07	0.62 ±0.15	0.63 ±0.16	0.66 ±0.19	0.64 ±0.22
GeoVex - 土地利用	3	2.97 ±1.43	0.14 ±0.05	0.66 ±0.15	0.51 ±0.17		0.70 ±0.05	0.71 ±0.23	0.72 ±0.22	0.71 ±0.24	0.62 ±0.25
GeoVex - OSM	1	3.84 ±1.56	0.14 ±0.05	0.60 ±0.12	0.46 ±0.12		0.76 ±0.06	0.64 ±0.1	0.65 ±0.1	0.63 ±0.12	0.76 ±0.22
GeoVex - 統合	1	2.87 ±1.26	0.14 ±0.05	0.57 ±0.13	0.44 ±0.15		0.75 ±0.06	0.71 ±0.13	0.70 ±0.11	0.70 ±0.16	0.65 ±0.24

Hex2Vec よりも森林や都市域など生態系や地域別の特徴の違いが小さくなったと考えられる。

下流タスク別の性能評価実験の結果からは、特に社会経済に関わる変数の予測において人工的地物の情報が多く含まれる OSM の情報を用いた場合の性能が高くなった。保全活動の実施の有無や人々の自然に対する価値の認識など、人間の活動や社会の状態との関わりが深いタスクに対しては、埋め込みベクトルを用いることが有効である可能性が示唆された。

5. 今後の展望

地理空間情報の表現学習は比較的新しい分野である。埋め込み手法そのものについては、1. 各埋め込みモデル別の適切な空間解像度の検討、2. 距離的に近いかどうかを判別する、社会・生態システムの機能の類似性を判別するなど、表現学習のタスクの検討、3. リモートセンシングとの連携した継続的なデータ作成、また、埋め込みベクトルの解釈可能性の向上に向けては、4. 埋め込みベクトルの軸の意味的解釈手法の開発、5. 埋め込みベクトルに対する各地物情報の寄与の評価、さらに、埋め込みベクトルを応用した地理空間情報解析のために、6. 文章や写真、生物情報など様々なデータとのマルチモーダル学習への発展が期待される。これまでに蓄積されてきた多様な地理空間情報を統合して解析することで、地域内の自然の多様な価値を認識・発見することにつながることを期待する。

謝辞

本研究成果は、大阪大学 D3 センターの大規模計算機システムを利用して得られたものです。また、環境省・(独) 環境再生保全機構の環境研究総合推進費 (JPMEERF23S12116) の支援を受けました。研究の構想・実施の段階では、国立情報学研究所の北本朝展教授、大阪大学工学研究科の松井孝典助教に貴重な議論の機会をいただきました。ここに謝意を表します。

参考文献

- (1) IPBES, IPBES secretariat, 620, (2022).
- (2) S. Woźniak, P. Szymański, Proc. of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, 61–71, (2021).
- (3) D. Donghi, A. Morvan, Proc. of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, 3–13, (2023).
- (4) Y. Gao, et al., Applied Sciences, **12**, 24, 12942, (2022).
- (5) M. Akasaka, et al., Integrative Observations and Assessments, 209–229, (2014).
- (6) 小川みふゆら, 保全生態学研究, **18**, 69-76, (2013).
- (7) H. Yamagiwa, et al., arXiv, (2023).
- (8) 芳賀智宏ら, 環境情報科学論文集, **38**, 141-147, (2024).
- (9) W. R. Tobler, Economic Geography, **46**, 234–240, (1970).