

Abstract: Remote sensing data is a globally and equally available data source and LULC (Land use and Land Cover) identified by remote sensing data is the representative product. Recently, LULC classification using machine learning technology gains attention. This study aimed to develop LULC classifiers using Landsat-8 OLI, Climate features, and GBDT (Gradient Boosting Decision Tree) algorithm. Seasonal composites of Landsat-8 OLI images and annual precipitation and mean annual temperature were used as radiation and climate features, respectively. The overall accuracy of 10 classes using only Landsat-8 OLI was 67.5 %, and using both Landsat-8 OLI and climate features was 69.2%. The classification performances of water and paddy land were high and forest types were low. Modifying pre-processing, updating machine learning algorithms, and introducing new features were proposed for future perspectives.

Keywords: land use/land cover (LULC), remote sensing, Landsat-8, machine learning, classification, GBDT

1. はじめに

リモートセンシングデータは全球で情報格差のないデータリソースであり、SDGs (Sustainable Development Goals) 指標のモニタリングと親和性が高い。SDGs のモニタリング対象の一つに土地利用土地被覆 (LULC: Land-Use / Land-Cover) があり、信頼性の高いアプローチの一つである機械学習を用いた LULC 識別が注目を集めている。そこで本研究では Landsat-8 衛星から得られたリモートセンシングデータと機械学習を用いて LULC 識別器を構築し、その精度を向上することを目的とする。

2. 分析方法

2. 1 分析対象と学習用・テスト用データセットの作成

2014~2016年の日本全域を対象に、水域・都市・水田・畑地・草地・落葉広葉樹 (DBF: Deciduous Broadleaf Forest)・落葉針葉樹 (DNF: Deciduous Needleleaf Forest)・常緑広葉樹 (EBF: Evergreen Broadleaf Forest)・常緑針葉樹 (ENF: Evergreen Needleleaf Forest)・裸地の10クラスへの分類を行った。

入力データとして、(1) JAXA 地球観測研究センターが提供する SACLA-J²⁾ の学習用教師データのリスト (N=34,088), (2) USGS が公開している Landsat-8 OLI (Operational Land Imager) の7バンドに Collection 1 Level-2 処理を行った 3,725 シーンのラスターデータ, (3) 年降水量・年平均気温のラスターデータ, (4) JAXA 地球観測研究センターが収集したテスト用教師データのリスト (N = 3,000) を取得した。Landsat-8 OLI データの Pixel Quality Assessment Band で図1のような雲などのノイズを除去し、欠損のない同じ季節かつ同じ地点の値で補完した。学習用・テスト用教師データの緯度経度の地点で (2) と (3) を 1×1 pixel で抽出して学習用データセット (N = 34,088) と、欠損値を除いたテスト用データセット (N = 2878) を作成した。

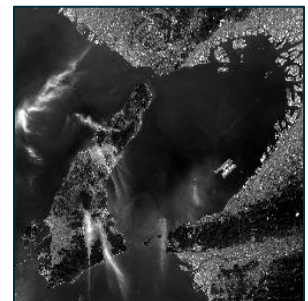


図1 Landsat-8 OLI

2. 2 二つの LULC 識別器の構築と汎化性能の評価

学習用・テスト用データセットに対して、GBDT (Gradient Boosting Decision Tree) の一種であり計算負荷と精度が両立する LightGBM 2.3.0 を使用して、二つの LULC 識別器を Python 3.6.5 上で実装した。LULC 識別器 α は入力データ (2) のみで学習し、LULC 識別器 β は入力データ (2) と (3) で学習させた。学習時にはデータをランダムに分割して 10×10 の入れ子式交差検証を行い、内側の交差検証ではベイズ最適化でハイパーパラメータを決定し、外側の交差検証では汎化性能を評価した。

	Prediction										正解	Recall
	水域	都市	水田	畑地	草地	DBF	DNF	EBF	ENF	裸地	データ数	(%)
Reference	水域	260	1	1		1		1	5	3	272	95.6
	都市	3	257	7	13	8		1		5	294	87.4
	水田		3	270	11	7	1	1		1	294	91.8
	畑地	1	7	31	200	24	7	3	8	3	290	69.0
	草地	1	1	12	19	210	27		8	4	284	73.9
	DBF	1		1	1	10	221	3	2	46	285	77.5
	DNF			1	4	3	135	88		66	297	29.6
	EBF	2	2	2	4	5	20		121	129	289	41.9
	ENF			1			11		10	272	294	92.5
	裸地	30	109	10	15	16	5		1	1	92	279
予測データ数	298	380	336	267	283	428	94	153	526	113	2878	
学習用データセット数	1714	3500	8775	6313	3807	3246	438	1242	4176	877	34088	
Precision (%)	87.2	67.6	80.4	74.9	74.2	51.6	93.6	79.1	51.7	81.4	OA	69.2

図2 LULC 識別器 β の 10 クラス混同行列

DBF = Deciduous Broadleaf Forest, DNF = Deciduous Needleleaf Forest, EBF = Evergreen Broadleaf Forest, ENF = Evergreen Needleleaf Forest

3. 結果・考察

LULC 識別器 α の OA (Overall Accuracy: 全予測正答率) は 67.5% , LULC 識別器 β の OA は 69.2% であり, Landsat-8 OLI データに加え, 年降水量・年平均気温を特徴量に追加することで精度は向上した. 識別器 β の混同行列を図2に示す. 水域や水田の精度は高かった一方で, 分類クラスごとの学習用データ数が不均一な森林に属するクラスの精度が低かった. 森林に属するクラスのうち学習用データセットが最も少ない DNF は, Precision (正予測の正答率), Recall (正に対する正答率) がそれぞれ 93.6% , 29.6% となった一方で, 学習用データセットが最も多い ENF では 51.7% , 92.5% となった. これは学習用データセットの多いクラスに偏って識別器が予測していることを示している. また, 図3は森林の季節での分類結果を示しており, 春 (1~3月) の分類性能が最も高かった. 一方で, 落葉樹を示す赤色と常緑樹を示す緑色の交点付近で誤分類を示す青色が多発していたため, 前処理の改善が必要である.

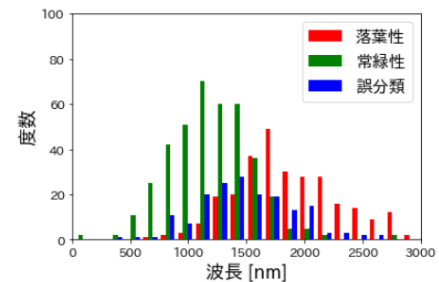


図3 1~3月のバンド6: 1560 - 1660nm

4. 今後の課題

汎化性能を向上させるために, 学習用データセットの作成プロセスで, 1) Landsat-8 OLI データの大気などのバックグラウンド値を考慮した前処理を行うこと, 2) より高解像度な光学衛星の Sentinel-2 や合成開口レーダなどの特徴量を追加することを検討する. また識別器では, 3) 分類クラスごとの学習用データセット数を調整する, 4) Cost-sensitive learning を行う, 5) CNN や Visual Transformer など画像識別型の深層学習を応用することが必要である. さらに, 作成した LULC プロダクトから, SDGs Indicator 2.4.1: 生産的で持続的な農業の下で行われる農業地域の割合, 11.3.1: 人口増加率と土地利用の変化, 15.1.1: 土地全体に対する森林の割合などの指標を評価するプロセスを開発することを今後の課題とする.

参考文献

- 1) A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data", GIScience & Remote Sensing, vol. 57, no. 1, pp. 1-20, 1月 2020, doi: 10.1080/15481603.2019.1650447.
- 2) 第一宇宙技術部門, "土地被覆図のための参照用地上データセット (SACLAJ)", Available: <https://www.sapc.jaxa.jp/use/rd/saclaj/>, アクセス日: 2021年1月31日