

0-030

アンサンブル学習型ニューラルネットワークを用いた衛星画像の水田抽出モデル Ensemble Learning of Artificial Neural Networks for Paddy Field Classification using Spatiotemporal Remote Sensing Data

岸田 和也† 山口 崇志† マッキン ケネス ジェームス† 永井 保夫†

Kazuya Kishida Takashi Yamaguchi Kenneth James Mackin Yasuo Nagai

1. はじめに

近年、衛星を用いたリモートセンシングによる水田の観察が重要視されている。観察される要因として、稲作は大量の水資源を必要とする農業の一つであり河川や湖沼といった灌漑施設から人工的に水を供給することにより水資源の枯渇が顕著化しつつあることがあげられる。さらに、水田は高い温室効果を示すメタンガスの主要な発生源であり、地球上で放出されるメタンガスのうちの20%が水田から放出されている。しかし、水田域の観察では地域や季節における環境の変化、データ量に起因する計算量の増大、雲によるノイズなどの問題などがあり、これまでの従来手法では有効な抽出モデルは得られていない。本研究では、過去の研究で比較的良好な結果の出ているニューラルネットワークを複数個組み合わせて精度向上が図れるアンサンブル学習を行った。アンサンブル学習手法は精度向上の為、重みつき多数決によってより良い分類器を生成するブースティング手法を適用した。

2. MODIS データ

2.1 MODIS

MODIS は 36 バンドの観測波長帯を持つ光学センサであり、NASA の Terra と Aqua の両方の衛星にそれぞれ搭載されている[3]。

表1 MODIS 要素

FOV	250m(2 bands),500m(5 bands), 1000m(29 bands)
観測幅	2300 km at 110°(±55°) from 705 km altitude
センサ幅	from 0.43 μm to 2.2 μm and ±45°scan
傾斜角	98 度
回帰日数	16 日
軌道	太陽同期準回帰軌道
量子化	12 ビット

2.2 評価基準データの作成

リモートセンシングの研究では、正規化植生指数、正規化水指数、正規化土壌指数などが情報抽出の為に用いられている。

植物の緑葉は青領域の波長を吸収し赤外線領域の波長を強く反射する特性がある。衛星リモートセンシングではこのような植物の特徴を生かし正規化植生指数(NDVI: Normalized Difference Vegetation Index)を用いている。可視バンド(R)と近赤外バンド(NIR)の値は、式(1)で求められる。

† 東京情報大学 総合情報学部 情報システム学科

正規化水指数(NDWI: Normalized Difference Water Index)は短波長赤外バンド(SWIR)と可視バンド(R)で式(2)で表わされる。

また正規化土壌指数(NDSI: Normalized Difference soil Index)は短波長赤外バンド(SWIR)と近赤外バンド(NIR)を用いて式(3)で示される。

$$NDVI = \frac{NIR - R}{NIR + R} \quad (1)$$

$$NDWI = \frac{R - SWIR}{R + SWIR} \quad (2)$$

$$NDSI = \frac{SWIR - NIR}{SWIR + NIR} \quad (3)$$

実験では、各正規化指数で扱われる可視バンド、近赤外バンド、短波長赤外バンドを、水田域における重要なバンドと仮定し 2002 年 1 月～12 月の千葉県北部地域の観測データを用いた。

水田域の教師データには土地利用 3 次メッシュデータを用いた。土地利用 3 次メッシュでは水田、畑、果樹園、森林、荒地、建物用地、幹線交通用地、湖沼、河川等の値が 1km ごとに格納されている。水田の値が最も高い地域を水田域とした。

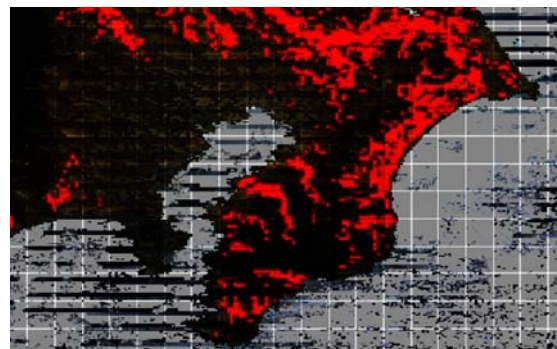


図1 2002年1月のMODISデータに3次メッシュ水田域のデータを重ね合わせたデータ表示

3. 提案手法

3.1. 集団学習

集団学習は、精度が高くない複数の弱学習器(本研究ではニューラルネット)を組み合わせることで精度を向上させる機械学習方法である[4]。異なる重み、あるいは異なるサンプルから単純なモデルを複数作成し、これらを多数決

などの方法で組み合わせることで、精度と汎用性を両立するモデルを作成することができる。

3.2. AdaBoost

AdaBoost は、集団学習でのブースティング手法の一つである。与えた教師付きデータを用いて学習を行い、その学習結果を踏まえて逐次に重みの調節を繰り返すことで複数の学習結果を求め、その結果を重み付き多数決で決めて精度を向上させる[5]。

AdaBoost 手法では 15 個のニューラルネットワークを学習器として用いて実験を行なった。

4. 実験

4.1. ニューラルネットワークの使用方法

学習器は過去の研究で有効な数値の出ているニューラルネットワークを用いた[2]。

ニューラルネットワークが学習しやすいように 0~1 の範囲に収まるように(4)式で正規化を行なった。

$$\text{正規化} = \frac{(X - X \text{ min})}{(X \text{ max} - X \text{ min})} \quad \begin{matrix} X = \text{正規化前の値} \\ X \text{ min} = \text{属性での最小値} \\ X \text{ max} = \text{属性での最大値} \end{matrix} \quad (4)$$

ニューラルネットワークは、階層型ニューラルネットワークを用い、学習方法には誤差逆伝播法(Back Propagation: BP)を使用する。入力データ数は 36 個とした。出力では水田と水田以外を分類させる。その為、教師信号に水田の場合に 1、水田以外の場合では 0 を出力させることにした。本モデルはニューロンを入力層に 36 ユニット、中間層に 36 ユニット、出力層に 2 ユニットで構成した。出力層の 2 ユニットは水田と水田以外のユニットに分け、出力した値の高い値で判別を行なった。また、学習係数(learning rate)は 0.3、収束速度(momentum)は 0.2 で学習を行った。

4.2. 実験詳細

全データ 20067 件から 10034 件を学習データとして用い、残りの 10033 件を評価用未学習データとした。まず、学習データを用いて 2000 回学習させ、その後評価用未学習データで実験を行なった。次に、学習セットを変更して 10 回実験を繰り返した。また、過去の論文との比較の為に同じデータセットを用いてニューラルネットワークだけでも実験を試みた。

4.3. 結果

表2 ニューラルネットワークと AdaBoost の正解率

	ニューラルネットワーク	Adaboost
平均正解率	86.2%	86.4%
正解が水田域の場合の正解率	73.7%	70.9%
正解が水田以外の場合の正解率	90.3%	91.6%

表2は、10 回実験を繰り返した平均正解率の結果を示しており AdaBoost の結果は 86.4%であり、ニューラルネットワークでの 86.2%と比べ 0.2%の向上が見られた。

また、図2はニューラルネットワークの学習における出力誤差を表わしている。学習回数は 1000 回に達している段階で出力誤差が収束している。

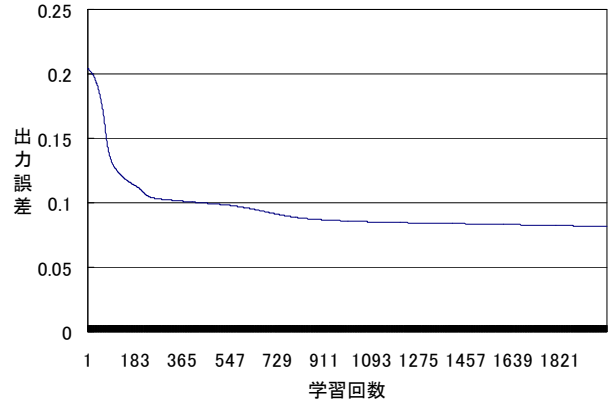


図2 ニューラルネットワークの学習における出力誤差

6. おわりに

ニューラルネットワークは、分類識別においてその識別関数学習の有効な手法である。パターン学習において広く利用されており水田抽出についてのニューラルネットワークを用いた研究はすでに行われている[1][2]。本研究ではニューラルネットワークを用いた AdaBoost 手法とニューラルネットワーク手法との比較を行なった。その結果 0.2%程度の正解率の向上が見られた。しかし、予想よりも正解率が上がらなかった。これは、元の MODIS データに含まれるノイズによってニューラルネットワークで正しく学習できずに正解率を下げていることが原因となっていると考えられる。今後、データの前処理や AdaBoost の重み付け方法、ニューラルネットワークの学習方法などを検討し、一層ノイズ多い問題に適した手法を提案する予定である。

参考文献

- [1]岸田 和也, 山口 崇志, マッキン ケネス J., 永井 保夫, “ニューラルネットワークを用いた TUR-BT 術後の膀胱癌の再発予測” 第 8 回情報技術フォーラム講演論文集(FIT2009), pp.537-538(2009)
- [2]T. Yamaguchi, K. Kishida, E. Nunohiro, J. G. Park, K. J. Mackin, K. Hara, K. Matsushita, and I. Harada, “Artificial Neural Networks Paddy Field Classifier using Spatiotemporal Remote Sensing Data” The 15th International Symposium on Artificial Life and Robotics, (2010)
- [3]森 一真, 山口 崇志, 朴 鍾杰, マッキン ケネス J. “Particle Swarm Optimization を用いた強調学習型 Artificial Neural Network による衛星画像からの水田抽出” 第 26 回ファジィシステムシンポジウム(2010)
- [4]元田 浩, 栗田 多喜男, 樋口 知之, 松本 雄治, 村田 昇, “パターン認識と機械学習 下” 株式会社シナノ(2008)
- [5]金森敬文, 畑埜晃平, 渡辺治, 小川英光, “ブースティング-学習アルゴリズムの設計技法-” 森北出版株式会社(2006)