

機械学習によるひまわり 10 号模擬観測データを用いた台風の識別
 Machine-Learning-Based Classification of Tropical Cyclones
 Using Himawari-10 Simulated Observation Data

○松本 涼太・榎本 剛
 ○Ryota MATSUMOTO, Takeshi ENOMOTO

The infrared sounder planned for launch aboard Himawari-10 in 2030 will provide observations across a greater number of spectral channels than conventional imagers, enabling more detailed retrievals of the atmospheric vertical structure. In this study, we applied various machine-learning techniques to simulated observation data designed to mimic those expected from the new sounder, with the aim of discriminating tropical cyclone and non-cyclone regions and classifying tropical cyclones by category. The results demonstrate that machine learning is highly effective for extracting and identifying tropical-cyclone features from high-dimensional satellite data, allowing dimensionality reduction and improved classification performance.

1. はじめに

衛星観測によるデータは、海上など観測点の少ない地点のデータをリアルタイムに取得できる点で非常に重要である。現在運用されているひまわり 8 号、9 号の後継機として、2030 年に打ち上げが予定されているひまわり 10 号には、今までのイメージと比較して、より多くの観測波長帯で観測することが可能になる赤外サウンダが搭載される予定である。これにより、大気の鉛直構造を把握することが可能になるため、線状降水帯や台風といった局地的な極端現象の予測や解明が期待される。一方、衛星観測により得られるデータは、多次元かつノイズが多く、データをそのまま解析に用いることは事実上難しく、次元削減など、何らかの方法で得られたデータから重要な情報のみを抽出する必要がある。本研究では、気象庁より提供された、ひまわり 10 号で観測されるデータを模した模擬観測データに対し、機械学習の手法を用い、台風と非台風領域との分類や、台風をカテゴリーごとに分類することを試みた。

2. 使用データと研究手法

ひまわり 10 号模擬観測データの存在する期間は、2020/06/09 – 2020/07/11、2021/06/09 – 2021/10/11、2021/11/09 – 2022/03/11、2022/06/09 – 2022/10/11 であり、この期間に存在する台風は 35

個である。次に、台風と比較する、対流活動が活発だが台風にまで発達しない領域（以下、非台風領域と称す）を、OLR（外向き長波放射量）の領域平均 $< 200 \text{ W/m}^2$ を基準として抽出した。それぞれ中心から $\pm 5^\circ$ の領域を画像として切り取り、128 pixel \times 128 pixel の画像として成形した。

台風と非台風を比較する 2 値分類では、台風のデータを増やすため、発生から 6 時間後までのものを含めて台風データセットに加えた。また、台風をカテゴリーごとに分類する多クラス分類では、模擬観測データの存在する期間内すべての台風を対象とし、データセットを構築した。

2 値分類、多クラス分類ともに機械学習の手法として XGBoost (Chen and Guestrin 2016) を採用した。この手法では、大規模データセットに対し高速かつ高精度で計算が可能であり、モデルの解釈可能性も高い。

それぞれの分類において、作成したデータセットのうち 80% を訓練データ、20% をテストデータとし、モデルの学習および分類性能の検証を行った。

3. 結果と考察

2 値分類の結果を表す混同行列を図 1 に示す。対角成分は予測と真値が一致している、非対角成分は予測と一致しなかったサンプルの数を示している。この結果から、本モデルは非常に高精度で台

風と非台風領域を分類できていることがわかる。また、多クラス分類の結果を表す混同行列を図2に示す。近接するカテゴリー同士では誤分類しているサンプルが存在するものの、短時間の学習によるモデルで良い性能が得られた。

画像分類においては、畳み込みニューラルネット

ワーク (CNN) が広く使用されているが、ハイパーカテゴリのチューニングに時間を要する上、豊富な計算資源を必要とする。本研究の結果は、衛星観測により得られるデータに対し、短時間かつGPUなどの計算資源を必要としない機械学習の手法が有効であることを示している。

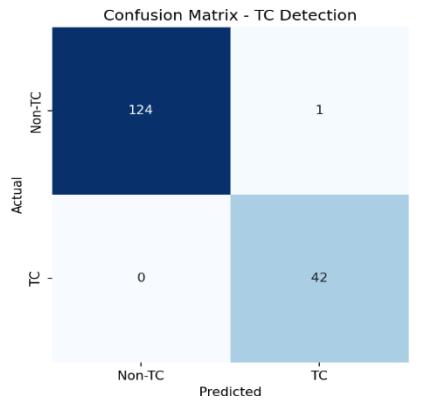


Fig.1 Confusion matrix for binary classification.

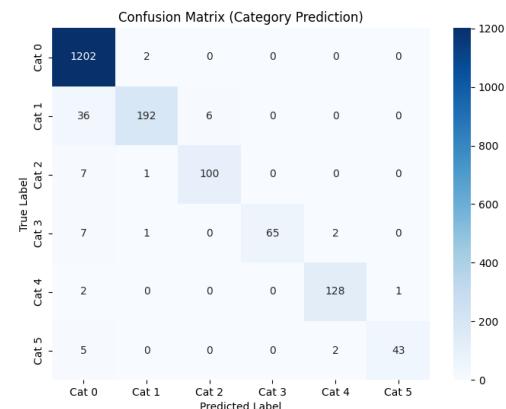


Fig.2 Confusion matrix for multi-class classification.