

## 気象変数の特徴量とした台風中心および対流活動域の識別 —XGBoostによる分類—

### Identification of Typhoon Centers and Convective Regions Using Meteorological Variables as Features: Supervised Classification with XGBoost

松本涼太<sup>(1)</sup>・榎本剛

Ryota MATSUMOTO and Takeshi ENOMOTO

(1) 京都大学大学院理学研究科

(1) Graduate School of Science, Kyoto University

#### Synopsis

Tropical cyclone genesis is a complex phenomenon influenced by various environmental factors, including temperature, surface pressure, humidity, and vertical wind shear. Due to its chaotic and nonlinear nature, accurate prediction remains a significant challenge. This study leverages machine learning techniques to identify key factors associated with tropical cyclones in the northeastern Pacific Ocean. Machine learning excels at extracting meaningful patterns and relationships from large datasets, enabling the discovery of previously unrecognized connections. We utilized a supervised learning approach, focusing on areas of active convection identified by Outgoing Longwave Radiation (OLR) data. To ensure physical interpretability and support domain-relevant insights, the modeling framework emphasizes transparency in feature contributions and decision boundaries. Specifically, we employed the XGBoost algorithm—a gradient boosting method well-suited for structured data analysis and feature-based classification tasks. By analyzing historical meteorological data, the model aims to determine the most influential parameters contributing to tropical cyclone formation, while maintaining a balance between predictive performance and scientific interpretability.

**キーワード:** 熱帯低気圧, 教師あり学習, 決定木, ブースティング, 解釈可能性

**Keywords:** Tropical cyclones, supervised learning, decision tree, boosting, interpretability

#### 1. はじめに

台風とは、熱帯低気圧が発達し、最大風速が一定以上となったもののことを指す。主に熱帯の海洋上で発生することが多く、その発生には気温、気圧、水蒸気量、水平風の鉛直シアなど種々の物理量が関与していると考えられる。発生後は、水平方向、鉛

直方向に複雑な構造を形成しながら北上し、中高緯度地域の国々に甚大な被害をもたらす。台風発生の予測には、その中心位置の正確な把握と、周辺の対流活動との識別が重要である。特に、対流活動が活発だが台風ではない領域（非台風領域と呼ぶ）との区別は、その後発達する台風が及ぼす災害リスクの評価という点で、熱帯気象学だけでなく、防災や保

険分野においても重要な意味を持つ。

台風の発生に関する先行研究として、のちの研究に大きな影響を与えたGray (1968)では、台風発生に好都合な条件として、環境場に関する6つの条件が提示されている。それらは次の6条件である。(i) 水平風の鉛直シアが小さい。(ii) 海面付近の水温26度以上である。(iii) 対流不安定である。(iv) 対流圏下層における低気圧性渦度が小さい。(v) 対流圏中層の相対湿度が高い。(vi) 赤道からある程度離れている。また、McBride and Zehr (1981)では、熱帯域での擾乱のうち、発達するもの (developing system) と発達しないもの (non-developing system) の違いについて、いくつかの特徴をあげている。たとえば、発達する擾乱では、低層の相対渦度が発達しないものの2倍以上である、台風が発生する場合、中心付近の鉛直シアが0に近い、発達する擾乱では、鉛直方向の温度勾配が顕著である、といった特徴である。

このようなパラメータが台風発生にどの程度関与しているのかを定量的に示す方法として、機械学習を用いることができる。機械学習とは、あるデータセットから規則などの特徴を抽出する手法で、近年では、計算機の発展により、特にヒトの神経組織を模倣したニューラルネットワークを用いた手法

(Amari 1967) が多くの研究分野で使用されている。機械学習における教師あり学習とは、ある入力 $x$ から $y$ を出力するための最適なモデルを学習するものである。

ニューラルネットワークを用いた手法の課題として、モデルの中身がブラックボックスであることが挙げられる。モデルがどのように判別を行っているかを確認するためには、モデルの精度を向上することだけでなく、解釈可能な手法を用いることが重要になる。解釈可能な教師あり学習により得たモデルの重みを確認することで、そのモデルが、どのような特徴量をもとに判別を行っているのかを知ることができる。

機械学習では、学習に使用するデータがモデルの精度に大きく影響するため、どのようなデータを入力に使用するのが重要になる。台風の特徴をより詳細にとらえるためには、台風中心付近と比較する領域として、衛星画像では台風であるか否かの判別が困難であるような、対流活動が活発な領域を選択することが望ましいと考えられる。

対流活動の活発さを示す指標として、外向き長波放射量(OLR)が広く用いられており、特に深い対流雲の存在を反映する。OLRの値が低いほど対流活動が活発であることを示すので、OLRは、台風やその前駆体の熱帯低気圧の抽出にも用いられる(Matsuoka et al. 2018)。本研究においては、台風中心

と比較する領域の選択にOLRを使用した。

本研究では、衛星データやデータ同化によって得られる物理量に対し、解釈可能な機械学習の手法を用い、台風の発生しやすい環境場を推定することを目的とした。機械学習モデル作成の教師データとして、台風中心位置およびOLRの値が低い非台風領域における、気温、絶対渦度、相対湿度および850hPa高度-200hPa高度間の鉛直シアを用いた。出力として、台風中心では"TC"、その他の領域では"NO"のラベルを与えた。これらの教師データをもとに、XGBoostによる学習によってモデルを作成し、学習には使用していないデータによってその精度を検証した。また、モデルの重みから、判別に重要な影響を及ぼしている特徴量を確認し、その気象学的な意味について考察した。

## 2. 使用したデータと解析方法

本研究では、北半球の太平洋上で発生する台風に着目するため、対象を東経120度～東経160度、および北緯2.5度～北緯20度の範囲に限定した。台風のデータとして、限定した範囲に発生した2018年～2022年までの台風を選択した。台風発生日時および中心付近の座標については、気象庁ベストトラックデータを使用した。該当領域における非台風領域として、第1節で述べたようにアメリカ海洋大気局(NOAA)の提供するOLRを使用し、10度四方の領域ごとの平均したOLRの値が250W/m<sup>2</sup>以下となる領域の中心を選択した。また、選択した非台風領域が、台風の前駆体あるいはすでに発生している台風領域と被ることを避けるため、前後3回以内に該当領域に台風が発生していない日時のデータを選択した。

選択した台風中心および非台風領域における物理量は、NCEP/NCAR再解析データの値を使用した。与える特徴量として、気温・絶対渦度(1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10hPa)、相対湿度(1000, 925, 850, 700, 600, 500, 400, 300 hPa)、850hPa高度-200hPa高度間の鉛直シアの計43個を選択した。

本研究では、教師あり学習の手法として、勾配ブースティング決定木: XGBoost (Chen et al. 2016)を採用した。この手法は、決定木を複数組み合わせたアンサンブル学習であり、解釈は容易だが精度が低い、決定木の課題を克服している。XGBoostでは、予測と真値とのずれ(コスト)を最小にするように、1つ目の決定木から順番に、1つ前の決定木での誤差を小さくするように学習を進めていく。この手法の利点は、大規模データセットに対しても高速で動作し、複数の決定木によって構築されていることからモデルの

解釈が可能な点である。XGBoostにより学習したモデルの重みは、決定木の分岐としてその物理量がどれほど使用されたか、すなわちその物理量がどれほど”TC”と”NO”の判別に影響を与えたかを示す。モデルの重みを見ることにより、どのような特徴量が台風領域と非台風領域の識別に重要であるかを確認した。

### 3. 結果

台風中心付近のデータ91点と、非台風領域のデータ144点を1つのデータセットとした。このうち、ランダムに振り分けた70%を学習用に用い、残り30%を学習には使用せずテスト用として用いた。機械学習の手法としてXGBoostを用い、学習したモデルをテスト用データで検証した結果が Fig. 1 である。予測ラベルと実際のラベルが一致しているもの(“TC”→”TC”, “No”→”No”)をそれぞれTrue Positive(TP), True Negative(TN)と表し、予測ラベルと実際のラベルが一致しないもの(“TC”→”No”, “No”→”TC”)をそれぞれFalse Negative(FN), False Positive(FP)と表す。モデルの精度(Accuracy: (TP+TN)/(TP+TN+FO+FN))は99.0%となり、高い結果が得られた。

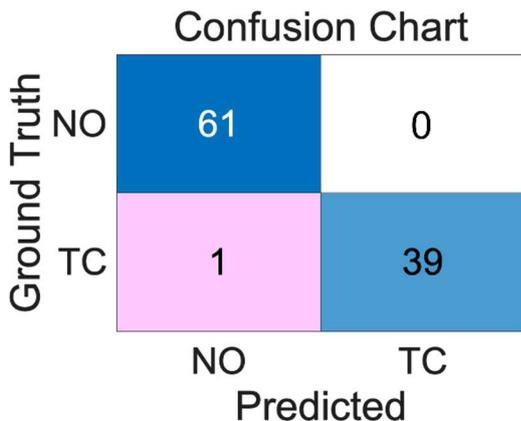


Fig. 1: This figure shows the confusion matrix for this analysis. The upper-left cell represents **True Negative (TN)**, where the prediction "No" matches the ground truth. The upper-right cell represents **False Negative (FN)**, where the prediction "TC" mismatches the ground truth. The bottom-left cell represents **False Positive (FP)**, where the prediction "No" mismatches the ground truth. Lastly, the bottom-right cell represents **True Positive (TP)**, where the prediction "TC" matches the ground truth.

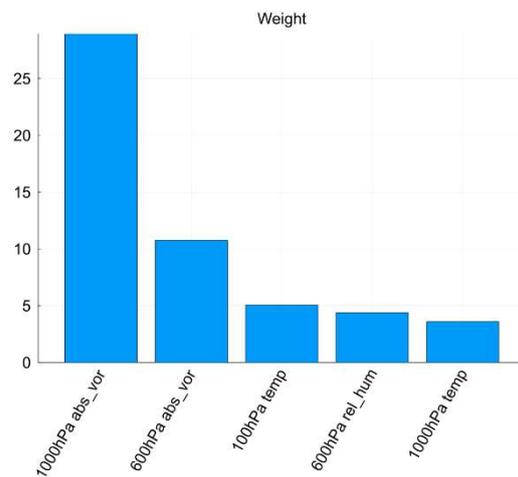


Fig. 2: The top five weights of the model's discriminator. The weights are normalized to sum to 100.

### 4. 考察

学習したモデルがどのように判別を行っているかを示す重みを、昇順に5つ挙げたものを Fig. 2 に示す。図の横軸は特徴量の名称、縦軸は各特徴量の重みの大きさを示す。ただし、特徴量全43個の重みの合計を100としている。これら5つの特徴量から、以下のように考えることができる。

- (i) 1000 hPaでの絶対湿度が判別に重要であることから、低層の絶対湿度の強さ
- (ii) 600 hPaでの絶対湿度が判別に重要であることから、(i)と併せて低層から中層にかけて鉛直方向に強い湿度を保つこと
- (iii) 100 hPaでの気温が判別に重要であることから、対流圏界面付近での気温が低いことによる対流不安定
- (iv) 600 hPaでの相対湿度が判別に重要であることから、中層での湿潤であること
- (v) 1000 hPaでの気温が判別に重要であることから、地表（海面）付近の温度が高いこと

この結果を第1節に述べた台風の特徴と比較して検討する。(i)は台風の最大接線風速が下層にあることを示している。(ii)は鉛直シアが弱いことに対応する。熱機関としての熱帯低気圧 (Emanuel 1986) において、上空における吹き出し域での冷却に対応していると考えられる。熱帯低気圧に伴う深い対流が発生するためには、中層が湿潤化し対流発生条件が整うことが必要である (Johnson et al. 1999)。(iv)は中層の湿潤化に感度があることを示している。地表面気温は海面水温を反映しているため、(v)は海面水温の重要性を示していると考えられる。すなわち5つの特徴量は、第1節で述べた

Gray (1968) ととも整合する.

## 5. 結論と課題

本研究では、発生直後の台風中心と台風活動が活発な非台風領域を比較し、XGBoostによる識別を試みた。精度の面では非常に高い結果が得られたが、サンプル数が少ないことや、台風中心のある1点のみを台風としてラベル付けしていることに不十分さがあると考えられる。今後の研究では、台風領域内の複数の点を使用するほか、重みづけするなどの方法をとることで、課題を解決できると考えられる。また、本研究では、気温、相対湿度、絶対渦度、鉛直シアといった、先行研究から台風発生に重要であると示されている物理量を特徴量として使用した。モデルの重みから、概ね台風の構造と一致するような結果が得られたが、既存の台風発生に関する研究との明確な違いを示すような結果は得られなかった。今後は、SSTや絶対渦度の時系列変化や、台風領域と非台風領域の差分を特徴量として加えることで、より詳細な結果が得られると考えられる。

## 謝 辞

本研究は、JSPS 科研費 21H04992 及び 23K26359 の支援を受けた。

## 参考文献

- Amari, S. (1967): A theory of adaptive pattern classifiers. *IEEE Trans. Electron. Comput.*, **EC-16**(3), pp. 299–307, doi:10.1109/PGEC.1967.264666.
- Emanuel, K. A. (1986): An air-sea interaction theory for tropical cyclones. Part I: steady-state maintenance. *J. Atmos. Sci.*, **43**(6), pp. 585–604, doi:10.1175/1520-0469(1986)043<0585: AASITF>2.0.CO;2.
- Gray, W. M. (1968): Global view of the origin of tropical disturbance and storms. *Mon. Wea. Rev.*, **96**, pp. 669–700, doi:10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2
- Johnson, R. H., T. M. Rickenbach, S. A. Rutledge, P. E. Ciesielski, and W. H. Schubert (1999): Trimodal characteristics of tropical convection. *J. Climate*, **12**(8), pp. 2397–2418, doi:10.1175/1520-0442(1999)012<2397: TCOTC>2.0.CO;2.
- McBride, J. L., and R. Zehr (1981): Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing systems. *J. Atmos. Sci.*, **38**, pp. 1132–1151, doi:10.1175/1520-0469(1981)038<1132:OAOTCF>2.0.CO;2
- Matsuoka, D., M. Nakano, D. Sugiyama, and S. Uchida (2018): Deep learning approach for detecting tropical cyclones and their precursors in the simulation by a cloud-resolving global nonhydrostatic atmospheric model. *Prog. Earth Planet. Sci.*, **5**, 44, doi:10.1186/s40645-018-0245-y.
- Chen, T., and C. Guestrin (2016): XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, San Francisco, CA, ACM, pp. 785–794, doi:10.1145/2939672.2939785.