

機械学習を用いたダム湖上流における有害藻類ブルームの発生予測モデルの開発と発生要因分析への応用

森雅也[†] 山城広周[†] 山本雄太[†] 森楓[†] 鈴木祥広^{††} 平岡透^{†††} 野中尋史[†]

[†]長岡技術科学大学

^{††}宮崎大学

^{†††}長崎県立大学

1 はじめに

近年、栄養素を多く含む生活排水や薬品などが川や湖、池へ流出することによる、有害藻類ブルームが問題となっている。有害藻類ブルームとは有害な藻類の異常発生であり、これは毒素や悪臭を排出することから、生態系のバランス崩壊に深く関与している [1]。また、ダム湖で発生した場合は下流河川にも影響する臭いや淡水赤潮など広範囲の公害の発生にも繋がるため深刻な環境問題となっている [2]。このことから、有害藻類ブルームの発生を予測するモデルや、発生の原因となる要因を分析/特定する必要性、重要性が示唆されている [3]。先行研究では、機械学習を用いた湖や沿岸海域での有害藻類ブルームの発生予測や発生要因に関する研究は盛んに行われているが、ダム湖やダム湖上流/下流を対象にした解析は現時点で行われていない [4]。そこで本稿では、川上が川下へ及ぼす影響を排除するために、ダム湖上流のみを対象とした機械学習による有害藻類ブルームの発生予測モデルの開発と発生要因の分析/考察を行う。具体的には、初めにダム湖上流の水質、環境、またこれらの短期間での分布の傾向を特徴量とし、それらによって特徴量空間を構築する。次に、特徴量空間評価関数を適用することで、クラス分類/異常検知問題に適した特徴量空間を獲得する。最後に、得られた空間に対しサポートベクターマシン (Support Vector Machine; SVM) で識別境界を引くことで、有害藻類ブルームの発生予測を行い、その時の特徴量を発生要因として分析/考察を行う。本稿では、

宮崎県の一ツ瀬ダムを対象とし、発生要因の解釈を容易にするために、特徴量空間の次元数を2次元とする。

2 実験概要

本稿では、一ツ瀬ダムの2003年1月から2019年10月までの3ヶ月おきのダム上流水質計測データとダム上流藻類発生量計測データを使用する。また発生予測は、特徴量エンジニアリングをしたその時点のダム上流水質計測データから、3ヶ月後のダム上流藻類発生量 (有害藻類ブルームを表す指標) が閾値を境にして正常か異常かを分類することで行う。

2.1 特徴量エンジニアリング

ダム上流水質計測データは、2003年1月から2019年7月までの3ヶ月おきの水温 (WT)、透明度 (CL)、水色 (WC)、溶存酸素量 (DO)、溶存酸素飽和率 (DSR)、伝導率 (Cond)、濁度 (Turb)、水素イオン濃度 (pH)、浮遊物質 (SS)、化学的酸素要求量 (COD)、リン酸 (Phos)、全リン (TP)、アンモニア (Ammono)、亜硝酸 (NSA)、硝酸 (NCA)、全窒素 (TN)、クロロフィル a (CA) の17種である。この17種に加え、窓幅を1年、スライド幅を3ヶ月とした各水質計測データの平均 (mean)、標準偏差 (std)、尖度 (kurt)、歪度 (skew) を、平均を0、分散を1で標準化したものを説明変数 (特徴量) とする。ただし、リン酸、アンモニア、クロロフィル a の尖度と歪度は欠損値を含んでいたため除外した。

ダム上流藻類発生量計測データは、2003年4月から2019年10月までの3ヶ月おきのラン藻、緑藻、珪藻、ミドリムシ藻類、黄色鞭毛藻類、渦鞭毛藻類、褐色鞭毛藻類の発生量である。これらの3ヶ月おきの藻類発生総和量が、平均である2573未満の場合を正常クラス、超えた場合を異常 (有害藻類ブルーム) クラスとし、これらのクラスラベルを目的変数とする。この時、正常クラスのデータ数は42、異常クラスのデータ数は21である。

Development of a Predictive Model and Application to Factor Analysis for Harmful Algal Blooms in Upstream Lake Using Machine Learning

Masaya Mori[†], Hirochika Yamashiro[†], Yuta Yamamoto[†], Kaede Mori[†], Yoshihiro Suzuki^{††}, Toru Hiraoka^{†††}, and Hirofumi Nonaka[†]

[†]Nagaoka University of Technology

1603-1, Kamitomioka, Nagaoka, Niigata, 940-2188 Japan

^{††}University of Miyazaki

1-1, Gakuen, Kibanadai-nishi, Miyazaki, 889-2192 Japan

^{†††}University of Nagasaki

Siebold Campus, 1-1-1, Manabino, Nagayo, Nishi-Sonogi, Nagasaki, 851-2195, Japan

s183368@stn.nagaokaut.ac.jp

表 1: 上位 20 種の特徴量空間の中で特に SVM による F 値が高い 3 種の特徴量空間の結果

特徴量	分類精度	F 値
(NCA-mean, DO-kurt)	.715	.676
(NCA-mean, DO-skew)	.717	.638
(TN-mean, WT-kurt)	.808	.628

2.2 予測モデル

本稿で扱うのは高次元小標本データなので、そのまま機械学習モデルを適用すると次元の呪いや過学習を起こす可能性がある。対策方法として、説明変数の次元削減が考えられるが、説明変数を合成変数に要約する方法では発生要因の解釈が困難になる。そこで今回は、各特徴量から 2 次元特徴量空間を構築することで次元の削減を行う。特徴量空間は、79 個の特徴量から構築可能な 2 次元空間として、 ${}_{79}C_2 = 3081$ 空間構築した。これらに、小標本データに対しロバストな特徴量空間評価関数である Minimum Reference Set(MRS)を適用することで、クラス分類/異常検知問題に適した特徴量空間を獲得した。獲得した上位 20 種までの空間に、カーネル関数を動径基底関数カーネル (Radial basis function kernel; RBF) とした非線形 SVM を適用することで、3ヶ月後のダム上流藻類発生量が正常か異常かを分類し予測を行なった。

3 結果と考察

本稿では、上位 20 種の特徴量空間に対して、5 回の層化 k 分割交差検証と SVM による分類精度/F 値を算出した。その中で、特に F 値の高い 3 つの特徴量空間の結果を表 1 に示す。また、最も F 値の高い特徴量空間を図 1 に示す。表 1 に着目すると、1 年間の硝酸 (NCA) の平均、1 年間の溶存酸素量 (DO) の尖度/歪度、1 年間の全窒素 (TN) の平均、1 年間の水温 (WT) の尖度が有害藻類ブルームの発生に関与している可能性が高いということが確認された。特に、1 年間の硝酸 (NCA) の平均と 1 年間の溶存酸素量 (DO) の尖度が重要であり、その際の分類精度は 71.5%、F 値は 67.6% ということが確認された。また、図 1 に着目すると、1 年間の硝酸 (NCA) の平均と 1 年間の溶存酸素量 (DO) の尖度が互いに高くなると、有害藻類ブルームが発生しやすくなるということが確認された。これは、1 年間の硝酸 (NCA) の平均が 130 以上、1 年間の溶存酸素量 (DO) の尖度が -1 以上の状態である。これらの結

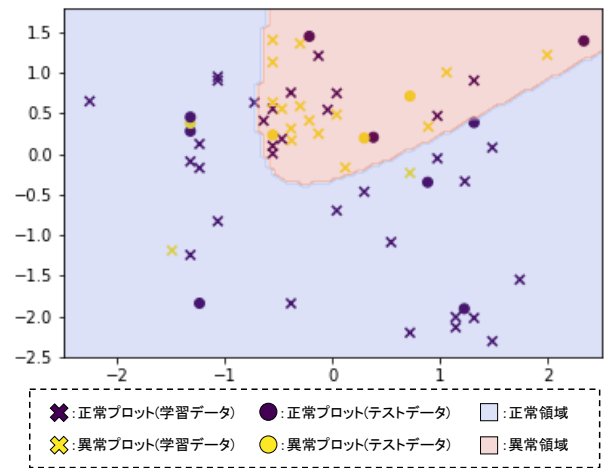


図 1: SVM による F 値が最も高い特徴量空間 (X 軸: NCA-mean, Y 軸: DO-kurt)

果から、有害藻類ブルームの発生には、1 年間での一定以上の栄養素と安定した酸素量が必要なのではないかと考えられる。

4 終わりに

本稿では、宮崎県一ツ瀬ダムの 3ヶ月おきのダム上流水質計測データから、3ヶ月後にダム上流で有害藻類ブルームが発生するか否かの予測モデルを構築し、その際の発生要因の分析/考察を行った。結果として、分類精度が 71.5%、F 値が 67.6%の予測モデルを構築することができた。また発生要因として、1 年間の硝酸 (NCA) の平均と 1 年間の溶存酸素量 (DO) の尖度が有害藻類ブルームの発生に深く関与しているという可能性が確認された。今後は、特徴量空間の高次元化による分類精度/F 値の向上を目指すと共に、実験的検証を行うことで、より深い検討を行なっていく。

参考文献

- [1] 松原賢, “有明海佐賀県海域における赤潮研究”, 佐賀県有明水産振興センター研究報告, vol.28, pp.129-133, 2017.
- [2] 朴木英治, 川村祐史, “有峰湖で 2019 年 7 月に観察された淡水赤潮とその臭気成分について”, 富山市科学博物館研究報告, vol.44, pp.27-33, 2020.
- [3] 金相暉, 藤井直紀, 濱田孝治, “機械学習を用いた沿岸海域における赤潮予測”, 土木学会論文集 B1(水工学), vol.75, no.2, pp.I-781-I-786, 2019.
- [4] J. Derot, H. Yajima, S. Jacquet, “Advances in forecasting harmful algal blooms using machine learning models: A case study with *Planktothrix rubescens* in Lake Geneva”, *Harmful Algae*, vol.99, 101906, 2020.