

伊万里湾における機械学習を用いた赤潮発生予測に関する基礎的研究

長崎大学工学部 学生会員 ○加来 太史
 長崎大学大学院工学研究科 正会員 鈴木 誠二
 長崎大学大学院工学研究科 正会員 田中 亘
 長崎大学大学院工学研究科 正会員 冨田 彰秀

1. 研究背景と目的

長崎県と佐賀県の県境に位置する伊万里湾(図 1)では、盛んに営まれる養殖業による多量の餌散布や陸域から流入する汚濁負荷の増大により、富栄養化が著しく進行し、赤潮発生が頻繁に確認されている。特に、2017年夏季には大規模な赤潮が発生し、養殖魚類に総額6億1千万円の甚大な被害が出た。赤潮の拡大防止や赤潮被害の軽減策などはあるものの、抜本的な赤潮の予防は困難な状況である。そこで、本研究では、機械学習を用いて予測モデルを作成し、気象および海象情報から赤潮発生予測の可能性について検討を行った。



図 1 伊万里湾

2. 赤潮発生予測の検討

2.1 対象海域の概要

図 2 に示す a~h 地点において赤潮の発生予測を行った。

2.2 研究手法の概要

本研究では、機械学習の分類という手法を用いて、赤潮発生予測の可能性を検討した。使用したデータは、気象庁から得られる日平均気温(°C)、日照時間(h)、降水量の日合計(mm)、日平均風速(m/s)、長崎水産試験場が計測している各定点における水温(°C)、塩分、さらに赤潮分布情報により確認できる各定点におけるカレンシア・ミキモトイの細胞数(cells/ml)のデータを用いた。対象期間は 2013 年~2018 年の 6 月~8 月とする。

2.3 予測モデルの概要

予測モデルの作成にあたり、平均気温、日照時間、日降水量、日平均風速、水温および塩分を機械学習における説明変数とし、赤潮発生の有無を目的変数とした。赤潮発生の有無に関しては、上記の対象期間において、カレンシア・ミキモトイの細胞数が 50cells/ml 以上の場合を赤潮が発生したと考え、赤潮が発生した場合の目的変数を 1、発生しなかった場合の目的変数を 0 とした。表 1 に示す通り、説明変数を赤潮の発生を考慮した日以前のデータを 3 パターンの期間に分けて機械学習を行なった。表 2 は各定点の赤潮発生件数と使用した総データ数である。

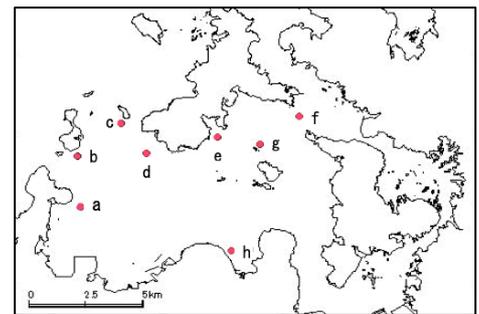


図 2 各定点

表 1 パターン

| | 説明変数 | 目的変数 |
|-------|----------|---------|
| パターン1 | 過去3日間の平均 | 赤潮発生の有無 |
| パターン2 | 過去5日間の平均 | 赤潮発生の有無 |
| パターン3 | 過去1週間の平均 | 赤潮発生の有無 |

表 2 赤潮発生件数

| 定点 | a | b | c | d | e | f | g | h |
|-----------|-----|-----|-----|-----|-----|----|----|-----|
| 使用した総データ数 | 107 | 102 | 105 | 106 | 104 | 82 | 99 | 103 |
| 赤潮発生件数 | 15 | 14 | 10 | 24 | 36 | 21 | 30 | 22 |

2.4 モデルの評価方法

式(1)~(4)に示す正解率, 精度, 検出率, F 値を求めた. 正解率(Accuracy)とは, 「赤潮発生と予測すべきデータを赤潮発生と予測し, 正常であると予測すべきデータを正常と予測している割合であり, 精度(Precision)とは, 「赤潮発生と予測されたもののうち実際に赤潮が発生した割合」であり, 検出率(Recall)とは, 「赤潮発生と予測すべき(実際に赤潮が発生)データを赤潮発生と予測できた割合」である. また F 値とは, 「精度と検出率をバランス良く持ち合わせているかを示す指標」であり, それぞれ 0~1 の数値で出力され, 0 の場合最も悪い評価, 1 の場合最も良い評価となる. それぞれの評価指標は混合行列(表 3)をもとに以下の式で求められる.

$$\text{正解率(Accuracy)} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{精度(Precision)} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{検出率(Recall)} = \frac{TP}{TP + FN} \quad (3)$$

$$F \text{ 値} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

ここに TP:実際に赤潮が発生しているケースをモデルが赤潮発生と判断した件数, TN:実際に赤潮が発生していないケースをモデルが正常であると判断した件数, FP:実際に赤潮が発生していないケースをモデルが誤って赤潮発生と判断した件数, FN:実際に赤潮が発生しているケースをモデルが誤って正常であると判断した件数である.

表 4 混合行列

| | | | |
|----|----|-------|----|
| | | 予測モデル | |
| | | 正常 | 赤潮 |
| 実際 | 正常 | TN | FP |
| | 赤潮 | FN | TP |

表 3 性能比較

| 定点 | c | e |
|-----|-------|-------|
| 正解率 | 0.834 | 0.755 |
| 精度 | 0.081 | 0.614 |
| 検出率 | 0.175 | 0.760 |
| F 値 | 0.094 | 0.671 |

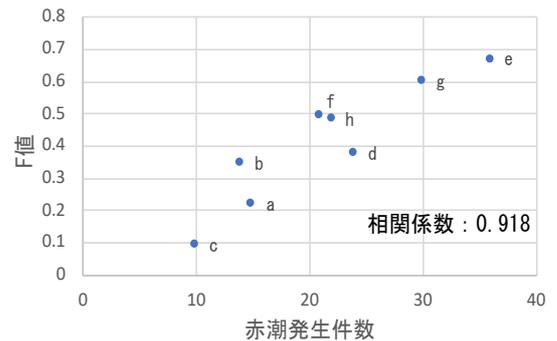


図 3 相関関係

3. 解析結果および考察

使用したデータの赤潮発生件数によって予測精度が大きく異なる結果となった. 赤潮発生件数が最も少ない定点 c の最も性能が良かったパターンのモデルと, 赤潮発生件数が最も多い定点 e の最も性能が良かったパターンのモデルで比較する. 定点 c のモデルの方が定点 e のモデルよりも正解率は高いが, それ以外の指標ではかなり低い結果となった. このことから, 赤潮発生件数が極端に少ないためモデルの学習が不十分で過度に正常と予測され, 正しく赤潮が予測できていない(TP がほとんど無い)ことが考えられる.

図 3 は, 赤潮発生件数と F 値が最も高かったパターンの各定点のモデルの F 値の関係性を示したものである. 正の相関関係があることがわかったので, 性能を向上させるために学習データを増やす必要があるといえる.

4. 参考文献

- 1) 沿岸海域水質・赤潮観測情報ポータルサイト, 沿岸海域水質・赤潮分布情報
http://akashiwow.jp/public/kaikuInit.php?qkaiku_id=1
- 2) 気象庁 各種データ・資料
<http://www.data.jma.go.jp/obd/stats/etrn/>