

# 気圧配置分類における類似度計算法の相互比較

\*佐藤拓人 (筑波大学生命環境科学研究科), 日下博幸 (筑波大学計算科学研究センター)

## 1. はじめに

気象予報や気象現象の解析の際, 気圧配置図を分類する機会は数多くある。データの大規模化を考慮すると, 分類手法の中でも, 計算機を用いた客観的な分類が今後ますます注目されるだろう。

客観的な分類では, 類似度の計算法の違いが分類精度に大きく影響する。気象分野では分類のための類似度として相関係数や二乗誤差, ユークリッド距離が類似度として使われてきた。最近では, Mo et al. (2014) が, 気象分野に限らず相関係数に関連した類似度計算法を実装し, 相互比較を行った。しかし, 画像ハッシュのように, 相関係数とはコンセプトの異なる類似度計算法に関する検討は未だ不十分である。また, 単一の典型的な教師データを設定した実験しか行っていない点も問題であろう。実際の分類では, ノイズを含むような典型的でない教師データを用いることも想定される。そのため, 教師データの違いによって精度がどの程度異なるのかも調査が必要だろう。

そこで本研究では, 4 つの類似度計算法(ユークリッド距離, 相関係数, structural similarity, average hash)を実装し, 複数の教師データを設定して精度を相互比較する。これにより, より実際のノイズを含む教師データの利用を想定した場合の, 気圧配置分類の精度向上に有用な情報を提供できると考えられる。

## 2. 手法とデータ

使用するデータは JRA-55(Kobayashi et al, 2016)の日本周辺の 10 年分の海面更正気圧である(14,612 枚)。対象とした気圧配置型は日本海低気圧型である。本研究では日本海上に低気圧の中心があるものを日本海低気圧型と定義し, 目視でラベリングを行った(839 枚)。

この日本海低気圧型のうち 1 つを教師データとし, その他のデータセット内の全てのデータとの類似度を計算する。類似度の高い順にデータセットを並べ替え, 先頭から特定の数データを抽出し, 取り出すことができた日本海低気圧型の全体に対する割合を抽出率とした。特に全ての日本海低気圧型(839 枚)を教師データとして抽出率の評価を行うことで, 教師データによってどの程度差があるかを調査した。

## 3. 結果

表 1 に, 各類似度計算法における精度を示す。Structural similarity は全ての教師データで評価した場合の平均的な抽出率も最高となった抽出率も, 他の類似度計算法より 5%以上高い。そのため, 教師データによらず, 他の手法より高精度になりやすいと言えるだろう。これに対して全ての教師データで評価した場合に最低となった抽出率を見ると, いずれの類似度計算法も 2%以下の精度であり, 計算法の差はほとんど見られなかった。この結果から, 客観的な分類において適切な教師を設定する必要があることが改めて確認できた。

## 4. 結論

本研究で対象とした類似度計算法の中では, Structural similarity が教師データによらず高精度で分類が可能であると考えられる。また, いずれの類似度計算法でも, 適切な教師を設定する必要があることを定量的に示した。

表 1 各類似度計算法の最高・最低・平均の抽出率(%)。抽出率は, 類似度の上位 839 位までに含まれる日本海低気圧型の枚数の割合とした。COR: 相関係数, EUC: ユークリッド距離, aHash: average hash, SSIM: structural similarity をそれぞれ表す。

Method	Maximum	Minimum	Mean
COR	31.9	1.5	11.6
EUC	27.7	2.0	10.6
aHash	29.3	0.6	11.2
SSIM	36.0	1.0	16.2

## 謝辞

本研究の一部は, (独) 環境再生保全機構の環境研究総合推進費 (2-1905) により実施された。

## 参考文献

- [1] Mo R. et al., 2014, *J. Hydrometeorol.*, **15**, 1862-1880.
- [2] Kobayashi C. et al., 2016, *J. geophys. Res. Atmos*, **121**, 1493-1510.