

自己組織化マップを用いた大気循環パターンのクラスタ解析

Clustering of Atmospheric Circulation Patterns Using Self-Organizing Maps

榎本剛

Takeshi ENOMOTO

Synopsis

The atmospheric circulation exhibits distinct patterns, such as west-high and east-low or south-high and north-low near Japan, typical during winter and summer, respectively, and bring associated weather and climate anomalies. These classification of the atmospheric circulation have been conducted subjectively or objectively. The empirical orthogonal functions (principal component) analysis can extract dominant patterns with small number of modes, but the relationship between different modes can be difficult to interpret due to the orthogonality. As an attempt to facilitate interpretation, a self-organizing Kohonen network is used to classify the atmospheric circulation patterns. After unsupervised learning of input data, similar patterns are mapped topologically close on the Kohonen network. The locations of positive and negative anomalies, gradients, and flow direction vary smoothly on the map. Atmospheric patterns near Japan were replicated using reanalysis for ten years.

キーワード: 異常天候, ニューラルネットワーク, 教師なし学習

Keywords: anomalous climate, neural network, unsupervised learning

1. はじめに

大気循環は、日本付近の西高東低の冬型や南高北低の夏型のような循環パターンを示し、それに伴い天候の偏差が生じる。大気循環のパターンの抽出は主観的または客観的に行われてきた。客観的な手法として広く用いられているのは、経験的直交関数 (Empirical Orthogonal Function, EOF) 解析である。EOF解析は統計学では主成分分析と呼ばれ、主要なパターンを少ないモードで抽出することができることが特徴である。一方、モード間の関係は直交性のために解釈が容易ではない。また、モードの線型結合で場を表現するため、非線型な関係を表現することはできない。

ニューラルネットワークは、多数の入力データから内部パラメタを反復的に調整することにより学習を行う。機械学習では、正解データを用いる教師あり学習と教師なし学習に分類される。教師あり学習では分類や回帰、教師なし学習はクラスタ解析や異常検知に用いられる。

ニューラルネットワークは、線型結合を用いているが微分可能な活性化関数を用いるため、非線型な関係を捉えることが可能である。

人工知能に応用されている深層学習では、入力と出力との間に複数の隠れ層を配置している。次元縮約には主成分分析を行う層が用いられることもある。

本研究では、隠れ層を持たないKohonen (1982)が提案した自己組織化ネットワーク (self-organizing maps, SOM) を用いて大気大循環の分類を試みる。SOMは節を2次元的に配置し、類似するパターン隣接するため、パターンの変遷に際する解釈がしやすいという特徴がある。SOMの手法については次節に述べる。

本研究では、欧州域の500 hPa面高度に対するクラスタ解析を行った先行研究 (Eckert et al. 1996) にならい、日本域においてクラスタ解析を行った。北半球冬季 (12月~1月) と夏季に分けて季節性を見るときともに、EOF解析との比較も行う。

2. データと解析方法

本節では自己組織化ネットワークのアルゴリズムについて簡潔に述べるとともに、利用したデータと実験の設定について記す。

2.1 自己組織化ネットワーク

SOMは教師なし学習の一つである。各格子点データは2次元ネットワークのユニットの一つ一つに対し、「刺激」として入力される。ユニットはそれぞれシナプス重みを持っている。初期値は小さな乱数で与える。ユニットの配置は矩形の他、左右に半分ずらした六角形の配置が用いられることもある。

類似の尺度は入力とユニットのシナプス重みとの距離で定義する。 p 番目の入力 Ψ^p とユニット i, j との距離は以下のように定義される。

$$D_{\text{RMS}}[w^{ij}, \Psi^p] = \sqrt{\sum_r [w_r^{ij} - \Psi_r^p]^2} \quad (1)$$

ここで r は物理空間での (1次元化された) 格子点番号である。

ネットワークのユニットの中で、この距離が最小となるユニットを選択ユニットとする。

選択ユニットからネットワークにおける距離が一定の距離を活動球とする。活動球においてユニットのシナプス重みを入力に近づける。

$$w_{k+1}^{ij} = w_k^{ij} + \tau_{(p)} [\Psi^p - w^{ij}] \quad (2)$$

活動球内のシナプス重みは相対的に大きく、活動球外の重みは小さくなることは、近隣のユニットと協調し、遠隔ユニットと競争していると考えることができる。この過程で類似したパターンが隣接する自己組織化が生じる。重みの変化量と活動球の大きさは学習回数とともに縮小させる。

大気循環を空間2次元の入力とした場合、個々のユニットのシナプス重みも空間2次元である。入力に最も類似するユニットが(1)で選択され、近隣の活動球内のユニットのシナプス重みが(2)により入力に近づけられていく。初期はランダムであるが、学習が進行するにつれてネットワーク上で類似したパターンを持つユニットが隣接するようになる。

2.2 データと実験設定

本研究で用いたデータとSOMの設定、比較のため行ったEOF解析について述べる。

(1) データ

米国環境予測センター (NCEP, National Centers for Environmental Prediction) / 国立大気研究センター (NCAR, National Center for Atmospheric Research)

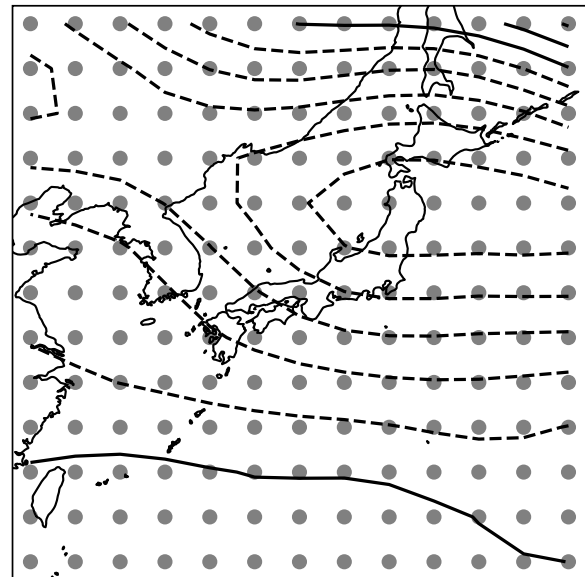


Fig. 1 Geopotential height anomalies on the 500 hPa surface at 00 UTC, 1 January 2010. Produced from NCEP/NCAR Reanalysis. Gray dots represents 13 × 13 data points in the 120°E–150°E and 20°N–50°N domain.

再解析 (Kalnay et al. 1996) から2010年1月1日から2019年12月31日までの10年間の500 hPa面高度について、6時間毎のデータから00及び12 UTC (7304時刻) を抽出した。このうち北半球冬季 (12, 1, 2月) の $t_{\text{max}} = 1802$ 時刻と夏季 (6, 7, 8月) の $t_{\text{max}} = 1840$ 時刻を用いて、それぞれの季節について学習を行った。解析の対象とした領域は日本付近の120°E–150°E, 20°N–50°Nで格子点数は13 × 13 = 169点である (Fig. 1)。

(2) 実験設定

ネットワークのユニット数は8 × 8、配置は矩形とし、乱数で重みを初期化した。先行研究とは異なり、特定の地点の月平均ではなく、各地点での月平均からの偏差を求め、偏差の標準偏差で規格化したものを入力として用いた。学習はランダムな順序で $p_{\text{max}} = t_{\text{max}} \times 5$ 回学習した。

学習率の初期値は $\tau_0 = 0.25$ で

$$\tau_{(p)} = \frac{\tau_0}{1 + p/p_{\text{max}}} \quad (3)$$

で小さくした。活動球の大きさの初期値 $\sigma_0 = 5.66$ で学習率同様の係数で縮小した。

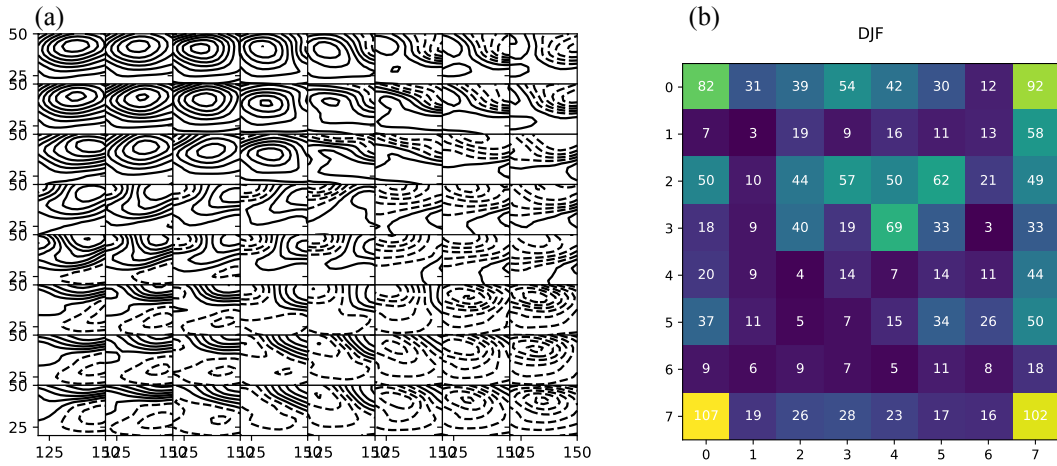


Fig. 2 Self-organizing maps for boreal winter. (a) Output synaptic weights and (b) frequencies.

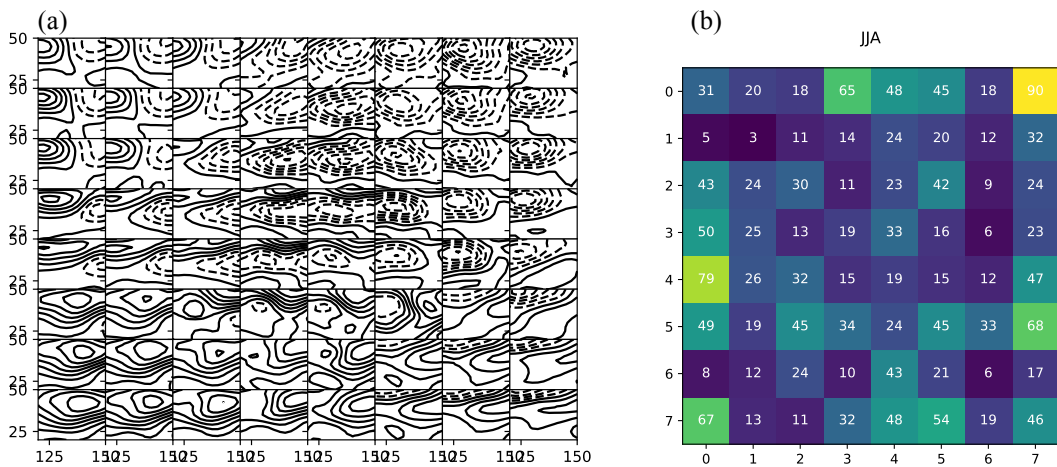


Fig. 3 As in Fig. 2 but for boreal summer.

(3) EOF解析

高度偏差を時間方向に規格化し、固有値解析を行った。固有値解析には特異値分解

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (4)$$

を利用した。Uは左特異ベクトル、Vは右特異ベクトルと呼ばれる。Σの*i, i*成分は特異値σ_{*i*}である。Uの列ベクトルは規格化された空間パターンを表し、Vの列ベクトルは規格化された主成分を表す。寄与率σ_{*i*}²/∑σ_{*i*}²で計算される。

SOMとEOFとの類似度は空間相関により計算した。ただし、EOFの符号は任意（逆符号もありうる）ことを考慮して、相関係数の絶対値が大きいものを最も類似しているユニットとして特定した。

3. 解析結果

北半球冬季及び夏季に対する学習結果をFig. 2及び3に示す。乱数を用いているため、結果は試行に依存するが、類似したパターンが試行毎にネットワーク上の異なる位置に現れる。ネットワーク上では隣接するユニットが類似しており、上下左右に遷移している。例えばFig. 2の左上 (0, 0) から下に向かって、南東側の負偏差により東風が領域の中央で強くなっていく。一方、右に向かって北東側に負偏差が現れ、北風が強化されている。遠く離れたユニット、例えばFig. 2の左上 (0, 0) と右下 (7, 7)は位相が反転している。

夏季 (Fig. 3a) は、冬季 (Fig. 2a) に比べてより水平スケールの小さな構造が重なっている。夏季には梅雨期が含まれるため、湿潤の効果により傾圧波のスケールが小さくなる (Tokioka 1971) ことを反映しているものと思われる。

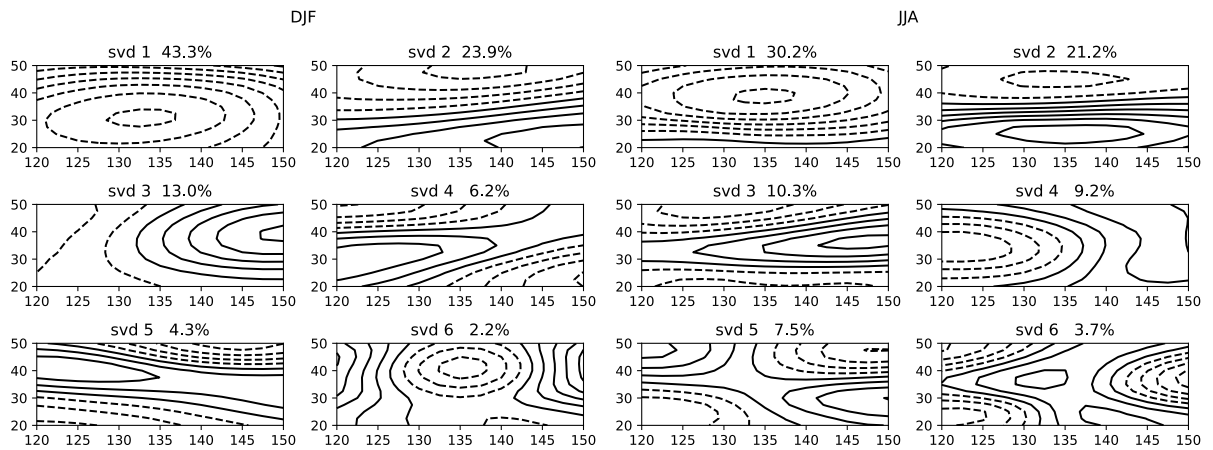


Fig. 4 Dominant EOF modes of 500 hPa geopotential height anomalies for boreal (a) winter and (b) summer.

Kohonen図は隅や辺が高頻度となる問題が指摘されている。実際冬季の四隅は頻度が高い (Fig. 2b)。一方、夏季の(0, 0)の頻度は31, (7, 7)は46で平均 (1804/64=28.75) を少し上回る程度にとどまる。頻度が0のような極端に低頻度なユニットは少ないことから、クラスタ解析は適切であると考えられる。

次にEOF解析を行い、SOMとの比較を行う。EOFの上位6モードをFig. 4に示す。卓越パターンを抽出するため、どちらも上位モードほど空間スケールの大きなものが選択されており、冬季と夏季とでスケールの差は見られない。EOFは互いに直交している。例えば冬季 (Fig. 4a) の第2モードの0線は第1モードの極値を通っている。

冬季及び夏季の第1モードはそれぞれ43.3, 30.2%の寄与率を示している。Kohonen図の最頻は冬季が(7, 0)の5.9%, 夏季が(0, 7)の4.9%である。Kohonen図は類似のパターンが近隣のユニットに現れるため、卓越したユニットが現れにくい。

冬季のEOFの1~3モードはそれぞれ(2, 4), (4, 7), (0, 7)に類似しており、相関係数はそれぞれ-0.88, 0.97, -0.89である。一方夏季のEOFの1~3モードはそれぞれ(7, 0), (5, 0), (3, 2)に類似しており、相関係数はそれぞれ-0.94, -0.91, -0.90である。

4. まとめと今後の課題

自己組織化Kohonenマップを用いて500 hPa高度を分類し、隣接するユニットが連続的に変化するネットワークを得ることができた。乱数で初期化するアルゴリズムであるため、学習のたびに異なる結果が得られる。結果の再現性については検討する必要がある。高頻度のユニットには主成分分析と一定の対応が見られた。

隅や辺に偏りやすい問題は、ユニットを上下左右に周期的とすることや球面上に準一様に配置されたユニットを用いることで解決できる。

今後は先行研究が行なったアンサンブル予報の信頼性の検証を行ったり、日本域や北半球など異なる領域や、月平均など異なる時間スケール、変数に対して適用したりすることが考えられる。深層学習のベースとして発展させていきたい。

謝辞

本研究はJSPS科研費19H05698の助成を受けた。

参考文献

- Eckert, P., D. Cattani, and J. Ambühl (1996): Classification of ensemble forecasts by means of an artificial neural network. *Meteor. Appl.*, **3**, 169–178.
- Kalnay, E., and Coauthors (1996): The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kohonen, T. (1982): Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, **43**, 59–69, <https://doi.org/10.1007/BF00337288>.
- Tokioka, T. (1971): A stability studies of medium-range disturbances with inclusion of convective effects. *J. Meteor. Soc. Japan*, **51**, pp. 1–10.

(論文受理日：2021年8月31日)