

異常性に着目した気象観測データの可視化

富岡 佑紀¹ 熊野 雅仁² 木村 昌弘²

1. はじめに

近年、世界の各地域で観測された気象データが、Web 空間上でオープン化され始めており、ビッグデータとして、データマイニングや知識発見の観点から、新たな活用法の探求や有益な情報の抽出に挑戦する研究が注目されつつある [1]。一方、情報可視化の研究においても、近年、気象データを対象として、空に占める雲の割合や、気温、気温の範囲、風速、雨が降る日の頻度などを、3次元CGによる四角柱の色、高さ、輝度や、複数の四角柱に関する配置上の規則性、密集度に当てはめ、多種情報を同時に世界地図上へ可視化することで、多種の気象データの特徴を一度に一望して比較分析し得る手法を提案している [2]。この可視化法は、気象に関する長期的傾向が捉えられ得るものの、近年懸念されている、短期的な気象の異常性を分析する上で、いつの時期にどれくらいの期間に渡ってどの程度の異常が起きているかなどを地域間で一望して比較分析するには不向きな面がある。我々は、気象データの可視化ではないが、撮影位置、撮影時間、撮影者などのメタ情報が付随した大量の写真群をうまく集約することで、空間上の人気撮影スポットを抽出し、さらに、各撮影スポットの時間軸上、人々が短期間により多くの撮影を行う異常性をバースト性の観点から捉えた旬シーズンを抽出し、各旬シーズンが、いつの時期にどれくらいの期間に渡ってどの程度のバースト性を示すかという多種情報を地図上に3D可視化する手法を提案している [3]。本研究では、旬シーズンの可視化に用いたアプローチを新たな課題にも適用できることを示すため、気象データにおける短期的な異常性を分析するための可視化問題に応用し、提案法の有効性を検証する。

2. 異常シーズンの抽出法

2.1 特徴量

本研究では、平年の値（長期の平均）から顕著に逸脱した気象データ上の値を異常値と考える。観測地域の集合を $\{R_1, \dots, R_K\}$ とする。正の整数 t に対し、地域 R_k における y 年、第 t 週目の気象観測データ一週間の平均値を $X_{k,y}(t)$ と

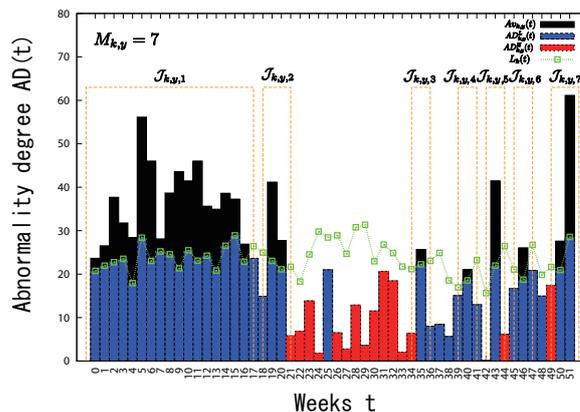


図1 1984年宮城県南部を R_k とする異常値の変動とシーズン

し、 R_k における $X_{k,y}(t)$ の30年平均値を平年値として $S_k(t)$ とする。ここで、平年値 $S_k(t)$ から外れた度合 (Abnormality degree) を

$$AD_{n,y}(t) = |X_{n,y}(t) - S_n(t)|$$

と定義する。ただし、 $AD_{k,y}(t)$ が $X_{k,y}(t) > S_k(t)$ を満たすとき $AD_{k,y}^H(t)$ 、 $X_{k,y}(t) < S_k(t)$ を満たすとき $AD_{k,y}^L(t)$ と表記して区別する場合がある。また、 t における $AD_{k,y}(t)$ の30年平均を $\mu_{AD_{k,y}}$ 、標準偏差を $\sigma_{AD_{k,y}}$ とし、 $L_k(t) = \mu_{AD_{k,y}} + \sigma_{AD_{k,y}}$ としたとき、地域 R_k における顕著に逸脱した異常値 (Anomaly value) を以下のように定義する。

$$Av_{k,y}(t) = \begin{cases} AD_{k,y}(t) - L_k(t) & (AD_{k,y}(t) > L_k(t)) \\ 0 & (\text{otherwise}) \end{cases}$$

図1における各 t の $AD_{k,y}(t)$ ビン上、赤を含むピンは $AD_{k,y}^H(t)$ 、青を含むピンは $AD_{k,y}^L(t)$ を表わしている。また、緑の点が $L_k(t)$ であり、黒い部分が $Av_{k,y}(t)$ である。

2.2 異常シーズンの抽出に基づく異常度

各地域 R_k の y 年における $Av_{k,y}(t)$ ($t = 1, \dots, 54$) に対し、 $M_{k,y}$ 個の異常シーズン $\{J_{k,y,1}, \dots, J_{k,y,M_{k,y}}\}$ を抽出する。ここに、各 $J_{k,y,m}$ は閉区間 $J_{k,y,m} = [t_{k,y,m}^{min}, t_{k,y,m}^{max}]$ である。 $m \leftarrow 0$ とし、 $t=1$ から $Av_{k,y}(t) > 0$ を満たす t を探索し、最初に見つけた t を $t_{k,y,m}^{min}$ とし、 $m \leftarrow m+1$ とする。 t が終端でなく、 $Av_{k,y}(t+1) > 0$ または $Av_{k,y}(t+2) > 0$ を満たしている間、 $t \leftarrow t+1$ を繰り返す、 $Av_{k,y}(t+1) > 0$ または $Av_{k,y}(t+2) > 0$ のいずれも成立しないとき、 t を $t_{k,y,m}^{max}$ とし、 $J_{k,y,m}$ を抽出

¹ 龍谷大学大学院理工学研究科電子情報学専攻

² 龍谷大学理工学部電子情報学科

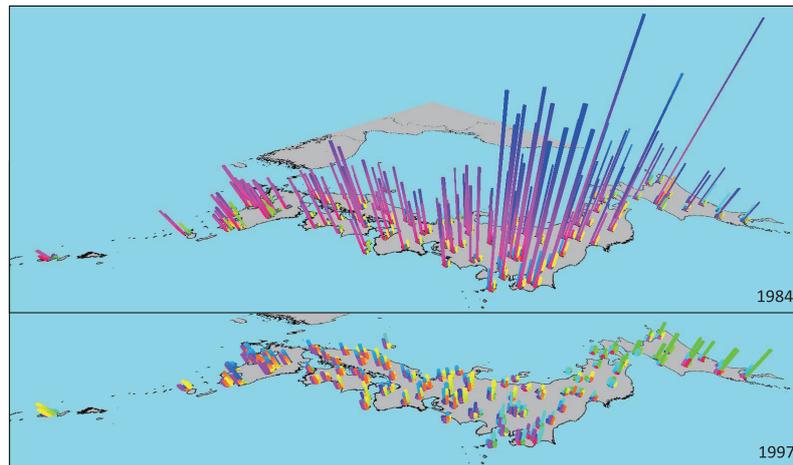


図3 1984年と1997年の気温に基づく異常シーズン可視化例

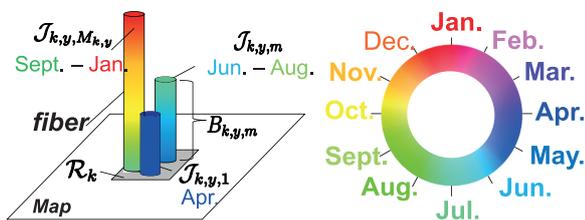


図2 各地域の異常可視化向けグラフ

する。以上の処理に基づき、 t が終端になるまで $J_{k,y,m}$ の探索を繰り返す。図1の橙色破線で囲った期間が抽出された $J_{k,y,m}$ の例である。ここで、各 $J_{k,y,m}$ の異常度を以下のように定義する。

$$B_{k,y,m} = \sum_{i=t_{k,y,m}^{\min}}^{t_{k,y,m}^{\max}} Av_{k,y}(i)$$

2.3 可視化法

本研究では、 y 年の R_k ごとに $M_{k,y}$ 個抽出された異常シーズンを図2左のように可視化する。図2左の一つの円柱を本研究ではファイバーと呼び、各 $J_{k,y,m}$ に対応させることで、各 R_k に何個の異常シーズンが抽出されたかを視認可能とする。また、各ファイバーの高さに $B_{k,y,m}$ を対応させることで、高いファイバーほど異常度の高さを視認可能とする。ファイバーの色は、図2右に示す色を割り当てる。この色は、1年間の各日にHSV色相環に基づく色を恣意的に割り当てたものであり、ファイバーの下端を $J_{k,y,m}$ の $t_{k,y,m}^{\min}$ 、上端を $t_{k,y,m}^{\max}$ とすることで、単色であるほど $J_{k,y,m}$ の期間が短く、多色であるほど長い期間であることが視認可能となる。

3. 実験

3.1 データセット

アメダスにおける平均気温および降水量データ(1981年~2010年)を用いた。またアメダス等の気象データは、

装置の休止や故障等により観測値が得られていない場合がある。そこで、一次細分区域に基づいて細分区域に属する観測地のデータを平均化して R_k とすることで、 $K = 143$ 個の観測地を可視化した。

3.2 可視化結果

図3が可視化結果である。紙面の都合上、代表的な結果として1984,1997年の気温のみの結果を示す。1997年では、多様な色のファイバーが多数存在し、複数の異常シーズンを持つ地域も多数存在することや、各 R_k の異常シーズンがいつの時期であるかが一望して同時に視認できることがわかる。一方、1984年では日本全域で1月や2月に極度の異常を示すことや、東日本では、ファイバーが赤から青まで変化している地域が多いため、5月まで異常が続いていることが視認できる。この年は、ラニーニャ現象が極めて強く影響した年であり、日本では5月上旬まで記録的な寒冬・寒春となった。提案法では、その影響がどの地域のどの時期のどの程度影響したかを一目で視認できることから提案法の有効性が示唆される。今後は、より詳細に視覚的分析を進め、より高度な特徴抽出法の構築や、可視化法の高度化を目指す予定である。

参考文献

- [1] J. H. Faghmous and V. Kumar: Spatio-Temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities. In W.W.Chu (ed.), *Data Mining and Knowledge Discovery For Big Data*, Studies in Big Data 1, pp.83-116, Springer-Verlag Berlin Heidelberg 2014.
- [2] C. G. Healey and B. M. Dennis: Interest driven navigation in visualization. *IEEE Transactions on Visualization and Computer Graphics*, Vol.18, No.10, pp.1744-1756, 2012.
- [3] S. Iwabuchi, M. Kumano, M. Koseki, K. Ono, M. Kimura: Visualizing attractive periods of popular photo spots using Flickr data, *Proc. of ACM SIGGRAPH'13 Posters*, Article112, 2013.