

気象データ可視化における二色塗り分け擬似カラー表示の 配色に関する考察

柴田 直道[†] 北 直樹[†] 斎藤 隆文[†]

東京農工大学 工学部情報工学科[‡]

1. はじめに

データの可視化は大規模なデータを分析するのに重要な手法である。大規模なデータの可視化において、全体像の把握のための情報と共に詳細な情報を提示することが一般的である。そのため、ユーザーは複数枚の画像をインタラクティブに切り替える必要がある。しかし、ユーザーにとって過度の画像の切り替えは理解を困難にすることがあるので、1枚の画像で全体像と詳細の両方を提示することが望ましい。その可視化手法の1つとして、二色塗り分け擬似カラー表示[1]がある。

各値を二色に塗り分ける事で、各点の値を正確に読み取れて、さらに値の変化がわかりやすいところが利点である。しかし、複数の条件を比較しようとした際に、同じ色を使用するとユーザーが混乱する恐れがある。そのため、本研究では複数のデータ種別を比較する際にも、ユーザーがそれぞれの条件の図であると認識できる様な配色を提案して考察する。

2. 二色塗り分け擬似カラー表示

二色塗り分け擬似カラー表示とは、時系列データを横軸に時間軸、縦軸に各時間ごとの値をとるような二次元画像として表す手法である。以下にアルゴリズムを示す。

まず与えられたデータ $f(x)$ (x : 単位時間)を元に、横の長さをデータの要素数、縦の長さを h とする長方形を作成する。

次にデータを m 個の色で塗り分けるために、 $m-1$ 等分する。 B を最初の区間の値、 A を区間の距離とすると、データがどの位置に分けられたか知るために、 $B + Ai$ ($i = 0, 1, \dots, m$)から i 番目にあることを求める。

$$B + Ai \leq f(x) < B + A(i + 1)$$

また、 $m + 1$ 個の色 $\{C_i(i = 0, 1, \dots, m)\}$ を用意する。

そして、データ $f(x)$ の各点に対してとなる i を求める。最後に色を塗り分けるために、 $(f(x) - (B + Ai)h/A)$ より上側を C_i で塗り、下側を C_{i+1} で塗る。

既存の手法では異なる種類のデータを比較した時に、色が重複していると分かりづらいといったことが起きる。そこで今回は5種類の気象データ(気温、降水量、日照時間、風速、風向)を使用して、それぞれのデータに適する配色を検討する。

3. 手法

それぞれの色分けを決めるために、まずそれぞれのデータの分布を以下表1から表5に示す。データの内訳は2020/1/1~2020/12/1までの47都道府県における5種類の条件を表したものである。また、この際に降水量以外の条件はデータを等間隔に分割したが、降水量だけは対数をとることで値を扱いやすくした。

表1 気温の度数分布

温度(°C)	-20~-10	-10~0	0~10
度数	2	265	4658
相対度数	1.0×10^{-4}	1.5×10^{-2}	2.7×10^{-2}
温度(°C)	10~20	20~30	30~40
度数	5729	6016	532
相対度数	3.3×10^{-2}	3.4×10^{-2}	3.0×10^{-2}

表2 降水量の度数分布

降水量(mm)	0~10	10~100	100~1000
度数	14764	2376	62
相対度数	8.5×10^{-1}	3.4×10^{-1}	3.0×10^{-2}

表3 日照時間の度数分布

日照時間(h)	0~4	4~8
度数	7185	4461
相対度数	4.2×10^{-1}	2.6×10^{-1}
日照時間(h)	8~12	12~16
度数	4789	767
相対度数	2.8×10^{-1}	4.4×10^{-2}

Study on the color scheme of two-tone pseudo coloring in visualization of metrological data

[†]Naomichi Shibata, Naoki Kita, Takafumi Saito

[‡]Tokyo University of Agriculture and Technology

表4 風速の度数分布

風速 (m/s)	0~4	4~8	8~12
度数	14466	2631	101
相対度数	8.4×10^{-1}	1.5×10^{-1}	2.8×10^{-3}
風速 (m/s)	12~16	16~20	20~24
度数	3	1	0
相対度数	1.7×10^{-4}	5.8×10^{-5}	0

表5 風向の度数分布

風向	N	E	S
度数	3942	3035	4262
相対度数	2.3×10^{-1}	1.8×10^{-1}	2.4×10^{-1}
風向	W		
度数	5963		
相対度数	3.4×10^{-1}		

二色塗り分け疑似カラー表示は連続的なデータを扱うのに適している。気温、日照時間の相対度数を見てみると全体的に分布しているので、連続的なデータであるといえる。それぞれの条件において異なる色を使用することで連続的であることを示す。また、風向においても同様のことがいえるが、風向は数値で表されていないので連続的ではない。

残りの2つである降水量、風速は共に相対度数が高いところがあるのでデータが密集していることから、非連続的なデータといえる。非連続的なデータで異なる色を用いると図が見にくくなるので、それぞれの条件において同種の色を用いる事で見やすくする。今回、比較のために同じ配色にしたものとそれぞれの分布から配色を変えたものを検討する。

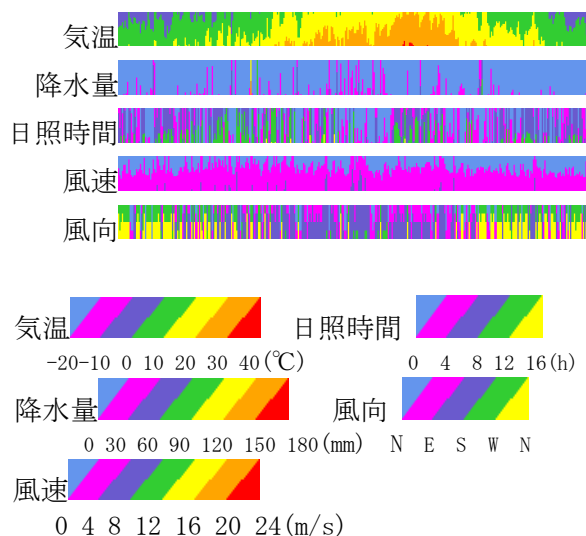


図1 同一配色による可視化

4. 結果

図1に同一配色による可視化、図2に提案手法による可視化を示す。

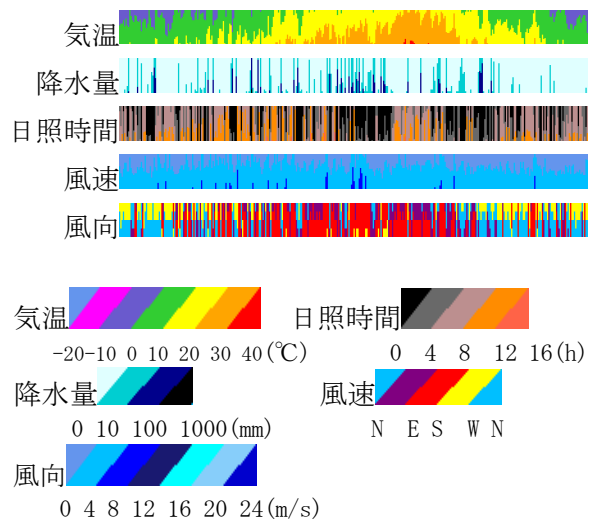


図2 提案手法による可視化

図1では同じ配色を使用しているの、同時に並べた時に異なるデータであると認識しづらい。一方で、図2では同じ配色ではないので別のデータであると認識することができる。

また、降水量は対数をとることでデータの幅を狭めたことで扱う色を少なくした。その結果、図1に比べて図2においては、より詳細にデータの推移を確認することができる。

5. おわりに

本稿では、複数条件における二色塗り分け疑似カラー表示の配色について提案と考察を行った。それぞれの条件の特徴を元に、二色塗り分けの特徴を生かした図を作成することができた。

今後は値の分布からだけでなく、時期によって生じる特徴的な値を考慮した配色も検討していく。

参考文献

[1] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya and T. Kaseda, "Two-tone pseudo coloring: compact visualization for one-dimensional data," IEEE Symposium on Information Visualization 2005, pp. 173-180. doi:10.1109/INFVIS.2005.1532144.