

7Y-02

二色塗り分けスパークラインを用いた COVID-19 感染データの可視化

本田 健悟[†]

斎藤 隆文[‡]

藤代 一成[†]

[†]慶應義塾大学 理工学部情報工学科

[‡]東京農工大学 工学部知能情報システム工学科

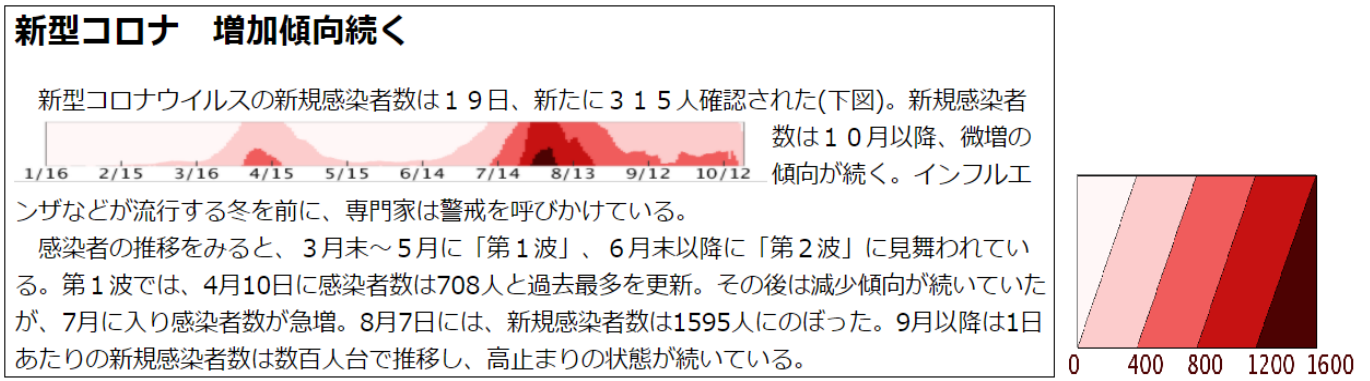


図 1: 時系列棒グラフに二色塗り分け法を適用したスパークライン. Web ブラウザ画面では, (a) 中のスパークラインにマウスをあてると, スパークラインの脇に (b) のレジェンドがポップアップされ, 値を読み取ることができる.

1 背景と目的

COVID-19 の感染データに関する可視化が連日メディアを中心に生成されており, Web 上や紙面で多くのスペースが割かれている. しかし, 棒グラフや折れ線グラフなどの可視化手法が多数を占め, スパークライン [1] などスペース効率のよい手法はほぼ使用されていない. そうしたなかで, 感染者数の推移などをわずかなスペースで細部まで読み取れるよう情報呈示ができれば, より多くの情報を従来のスペースで呈示可能になる. 例えば, 全国だけでなく, 地域や都道府県ごとの感染者数の可視化も従来のスペースで表示できる.

本研究では, 図 1 に示すように, 文章を見ながら図の概略を瞬時に理解することができるスパークラインに, 二色塗り分け法を適用させた可視化手法を提案する. 二色塗り分け法は, 折れ線グラフなどに比べ, 少ないスペースで情報を伝えることや値の変化の速度を即座に読み取れる. これをスパークラインに適用することで, 文章と図を同時に見ながら図の概略がわかるだけでなく, 詳細もすぐに読み取ることができるような可視化手法を実現する.

2 関連研究

スパークラインは Tufte[1] によって提案された, 任意の可視化を単語サイズでテキストに埋め込む手法であり, 図 1 のように使用されている. Microsoft の Excel やその他のソフトウェアにも実装され

ており, スプレッドシートや表で一般的に使用されている. 読者は文章と図を同時に見ることができるため, 内容への理解が容易になる. しかし, スパークラインは少ないスペースで図の概要を表示することに重点をおいているため, 元のグラフに比べて, 詳細な値が読み取れないという欠点がある.

二色塗り分け法は, 斎藤ら [2] が提案した手法で, 棒グラフや折れ線グラフなど 1 次元データを値域の区間ごとに水平に分割してそれぞれに異なる色を塗り, すべての区間の長方形領域を重ねることで, 変化量を表す波形を残したまま疑似カラー表示する手法である. 塗り分けた二色の比率をレジェンドの比率と照らし合わせることで, 個々のデータを正確に読み取ることができる. さらに, 増加や減少の速度を色の境界の傾きから読み取れる. そのため, 通常の可視化手法に比べ, スペース効率が高く, 限られた領域で可視化を表示する際に有効な手法である. 本研究は, スパークラインに二色塗り分けを適用することで, 両者の長所を活かした手法を提案する.

3 手法

本節では提案手法の手順を詳しく説明する.

3.1 スパークライン

スパークラインは, テキストに埋め込まれた単語サイズの可視化であり, テキストと同じ高さである 1 行で表示す

Visualizing COVID-19 Infection Data with Two-tone Pseudo Color-coded Sparklines

Kengo Honda[†], Takafumi Saito[‡], Issei Fujishiro[†]

[†]Department of Information and Computer Science, Keio University

[‡]Department of Electrical Engineering and Computer Science, Tokyo University of Agriculture and Technology



図 2: 時系列棒グラフを二色塗り分けスパークラインにした例. (a) に比べ、平滑化した (b) の方が値を読み取りやすい. また、(c) のように色の個数を減らすことで値の増加や減少の速度がよりはっきりわかる. (d) のようなグレイスケールテキストチャ表示でも値を読み取れる.

ることが多い. しかし、1 行では図の詳細を読み取ることが難しいため、本稿では可視化の詳細と概要の両方を読み取ることに重点をおき、2 行で表示する方針を採用した.

また、レジェンドは、スパークラインにマウスをあてることで表示されるようにした. レジェンドの大きさは、レジェンドの数値がテキスト 1 文字と同程度になるようにし、図 1 (b) のように、本文をなるべく遮らないように適当なサイズを選択した.

3.2 二色塗り分け法

二色塗り分け法において、何種類の色で塗り分けるか決めることは、値の読み取り精度にかかわる重要な問題である. しかし、『分かりやすい結果を得るために適切な色を配置することは難しい問題』[2] である. これは、カラーデザイン自体の難しさに加え、色を増やせば可読性は増すが視認性は低くなるというように、可視化する際に重要な両者がトレードオフの関係にあるためである. 色数を決める際には、表示媒体や対象年齢を想定し、最も伝えたい点を熟考する必要がある. 本稿では、Web 版の新聞記事を用いて幅広い年代の読者を想定する. 5 種類の色を用いて 0 から 1,600 までの値を区間ごとに徐々に色が濃くなるように塗り分け、各長方形領域を重ね合わせることで、図 1 のような二色塗り分け法を実装した. このように、棒グラフの波形を残すことで、情報量を減らすことなくスペース効率を高めている.

4 実行例

スパークラインの作成には、厚生労働省のオープンデータ [3] にある陽性者数の csv データを使用した. 2020 年 1 月 16 日から同年 10 月 19 日までの PCR 検査新規陽性者数を本手法で可視化した結果を図 1 に示す. 各値は 1 週間の

平均をとって平滑化した. スパークラインにマウスをあてることで、レジェンドがポップアップされ、図の概略だけでなく、値を即座に読み取ることができる. 通常のスパークラインと同じ大きさのとき、本手法のほうが効果的に情報を伝えていることがわかる.

二色塗り分け法は、高周波成分を含むデータの表示に不向きなため、スパークライン上で個々のデータが読み取れるよう、平滑化など予め元データを適切に処理することが重要である. 特に、日本の COVID-19 新規感染者数は平日と休日で大きな差があるため、本研究では、1 週間の平均をとって平滑化した. 図 2 (a) に平滑化していないスパークラインを、図 2 (b) に 1 週間の平均をとって平滑化したスパークラインを示す. 平滑化することで、感染者数の増加や減少の速度が読み取りやすくなっていることがわかる.

さらに、色数を減らすことで、値の増加や減少の速度が読み取りやすくなる. これは、個々の値が読み取りにくくなることとのトレードオフである. 図 2 (c) に 3 種類の色を用いて塗り分けたスパークラインを示す. 5 種類で塗り分けたときに比べ、値の増加や減少の速度が読み取りやすくなり、感染者数の推移がよりはっきり見れる. 平滑化し、色数を減らすことで



のように 1 行での表示も容易になる. また、色弱の方や白黒の紙面での表示を考慮して、グレイスケールテキストチャでの表現も実装した. 図 2 (d) に本手法をテキストチャで実装したスパークラインを示す. カラー表示のときと同様に、値の読み取りが可能であることがわかる.

5 結論と今後の課題

本稿では、二色塗り分けスパークラインを提案した. 本手法は、詳細と概要を同時に読み取ることができ、スペース効率のよい可視化手法である.

今後の課題としては、適切な色やテキストチャ、サイズの追究が挙げられる. 本研究で採用した色やテキストチャ、サイズは、人間の認知機能を十分に考慮できていない. 横軸の数値に関しても、最適な表示方法を追究したい. さらに、評価実験の実施が挙げられる. 既存の可視化手法と比べることで、実際に本手法が有効であるか確かめたい.

謝辞

本研究の一部は、令和 2 年度科研費基盤研究 (A) 17H00737 の支援により実施された.

参考文献

- [1] E. R. Tufte: *Beautiful Evidence*, Cheshire, Conn. : Graphics Press, 2006.
- [2] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya and T. Kaseda: "Two-tone pseudo coloring: compact visualization for one-dimensional data," in *Proc. IEEE Symposium on Information Visualization*, Minneapolis, October 2005, pp. 173–180.
- [3] <https://www.mhlw.go.jp/stf/covid-19/open-data.html>, 最終アクセス日: 2021 年 1 月 8 日.