

統計グラフ群の Small Multiples 表現による 地理空間データ可視化

佐藤 達哉[†] 大森 匡^{††} 新谷 隆彦^{††} 藤田 秀之^{††}

[†] 電気通信大学情報理工学域 I 類

^{††} 電気通信大学大学院情報理工学域研究科

1. はじめに

大規模データ処理の主要技術の1つとしてデータ可視化の研究が盛んである[1]. 本研究では, Small Multiples という小さなグラフを多数並べる可視化手法を対象とし, 地理空間データに適用した際の課題を解決する, 新しい可視化手法を提案する.

2. 課題

予備実験として, 地理空間データを, 異なる複数のサイズのグリッドによる領域分割を用いた Small Multiples 形式で可視化し(例を図1に示す), 以下のとおり主な課題をまとめた.

- 領域数が多いと読み取りが困難
- データ値が非常に小さくグラフが読み取れない領域が多数生じる
- 分割領域のサイズを大きくすると, サイズが小さい時に現れた領域内のデータの特徴や傾向が消失する可能性がある

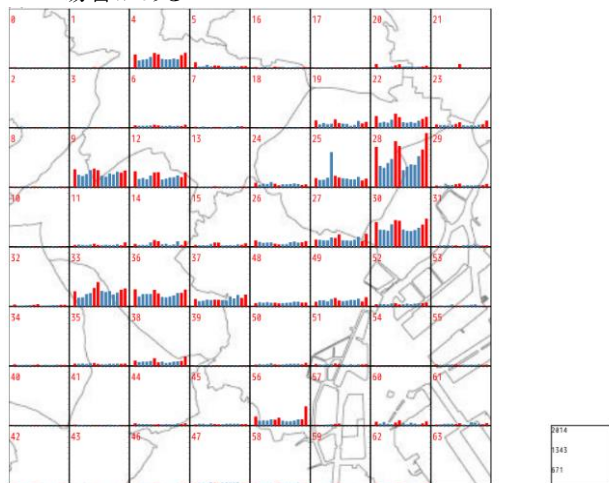


図1. Small Multiples による初等的な可視化.

3. 提案手法

本研究では, 以下の手順で可視化する.

1. グリッドで領域分割しデータを集約
2. 隣接するすべての2領域 $r_i, r_j \in \text{全域}R$ について結合適合性 $s(r_i, r_j)$ を算出
3. 結合適合性最大の2領域を1領域に結合(データを集約)
4. 軸スケールを計算
5. 手順2~4を一定回数繰り返し
6. 各分割領域内に統計グラフを描画

結合適合性 $s(r_i, r_j)$ は, 3種類の指標の線形和とした. ここで用いた指標は, データ遷移の類似度, 正方形に対する充填率, グラフの見かけ上の高さに基づく指標として提案した3種類である. 軸スケールは全域 R を用いて以下の式で求める.

$$\max_{r_i \in R} (r_i \text{ のデータ値の最大値} / r_i \text{ の} Y \text{ 軸方向ピクセル数})$$

4. 評価

各指標の使用/不使用の全ての組み合わせにおいて, 全域 R の領域分割状態を, (1)領域分割がデータ値の遷移の類似性を反映しているか, (2)各分割領域が正方形にどれだけ近いか, (3)各領域のグラフがどの程度の高さかの3つの観点から評価し, 2領域間の結合適合性の有用性を評価した. 結合適合性の指標を全て利用し, 可視化すると図2になる.

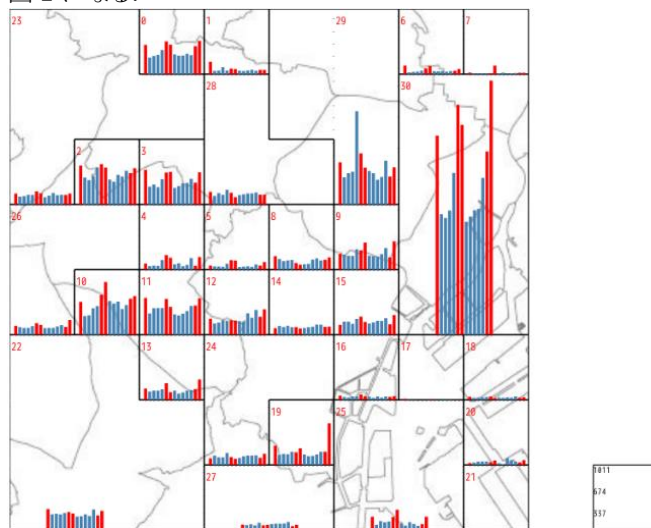


図2. データ遷移の類似度, 充填率, 見かけ上の高さを利用評価より, 各結合適合性は目的とした効果を持つことが確認でき, 各指標の重みが暫定的(1:1:1)でも適合性を全て利用すると良い可視化結果が得られた. ただし, 一部のデータ値が極端に少ない領域でグラフが見えない問題が残った.

5. おわりに

本研究は, 地理空間データの Small Multiples 形式での可視化において, 分割領域内のデータの類似性と, 可視化結果の可読性に着目した手法を提案して評価した.

参考文献

- [1] Meulemans et al., Small Multiples with Gaps, IEEE Trans. Visualization and Computer Graphics, 23(1), 381-390.