

連続面積カルトグラム作成アルゴリズムの高速化と大規模問題への応用

井上 亮, 北浦一輝, 清水英範

An Improved continuous area cartogram construction algorithm and its applications to large-scale problems

Ryo Inoue, Kazuki Kitaura, Eihan Shimizu

Abstract: We have proposed a simple algorithm for continuous area cartogram construction, which is one of the visualization techniques for spatial information. However, when input dataset is large, which means that the number of regions on a cartogram is large and /or the regions have detailed shapes, the calculation amount becomes enormous and the algorithm lack in feasibility. In this paper, we improve the applicability of area cartogram construction algorithm through the reduction of calculation applying the numerical calculation techniques, and also reducing the number of regions and the number vertices to input the algorithm simultaneously. Then we check the applicability of the improved algorithm by the application to the large-scale problem.

Keywords: 連続面積カルトグラム (continuous area cartogram), 視覚化 (visualization), アルゴリズム (algorithm), 高速化 (speed-up)

1. はじめに

近年、情報技術の発展や情報公開の取組等により、多種多様な統計データを容易に利活用ができるような環境が整ってきてている。筆者らは、統計データの有効活用を補助する一手段として統計データ視覚化に着目し、カルトグラムの作成手法を提案してきた。

カルトグラムの一種である連続面積カルトグラムは、地図上の面積で統計データの大小を表現するよう地図を改めて変形させることで統計データの空間分布を表現するものである。これまでに連続面積カルトグラムに対して、作成手法を提案した(井上・清

井上 : 〒113-8656 東京都文京区本郷 7-3-1

東京大学 大学院工学系研究科 社会基盤学専攻

Tel: 03-5841-6129, Email: rinoue@civil.t.u-tokyo.ac.jp

水, 2005)が、全国の市区町村人口データに適用したことろ、大型計算機を用いて 10 時間を超える計算時間を要し、大規模な問題に対して実用的に乏しいことが発覚した。

そこで本研究では、既存作成手法を基に計算手法を改良して計算量を削減し、大規模問題への応用が可能な手法を提案する。

なお、本論文では、連続面積カルトグラムを以後カルトグラムと記載する。

2. 既存のカルトグラム作成手法

井上・清水(2005)のカルトグラム作成手法について概要を示し、問題点を指摘する。

この作成手法は、地域形状を三角網分割しデータ

を各三角形に按分することを前提にしている。三角形の面積を按分されたデータに合わせる問題を最小二乗問題で表し、カルトグラムと地理的図上の地域形状を可能な限り近づけるため三角網辺の方位角変化を抑える正則化項を導入し定式化している。

作成手法は、頂点 i, j, k からなる三角形 t_{ijk} のカルトグラム上の面積 A_{ijk} をデータ D_{ijk} に合わせる目的関数と、頂点 m, n を結ぶ辺 e_{mn} の地理的図上の座北方位角 θ_{mn}^G をカルトグラム上の座北方位角を θ_{mn}^C に近づけて三角網の辺の方位角変化を抑制する正則化項からなる。三角網を構成する三角形の集合 T 、三角網の辺の集合を E 、正則化項の重みパラメータを μ とすると、式(1)で表される。

$$\min \left[\sum_{t_{ijk} \in T} \left(1 - \frac{A_{ijk}}{D_{ijk}} \right)^2 + \mu \sum_{e_{mn} \in E} (\theta_{mn}^C - \theta_{mn}^G)^2 \right] \quad (1)$$

ここで、頂点 i のカルトグラム上の座標 (x_i, y_i) 、地理的地点配置上の座標 (x_i^G, y_i^G) を用い、更にカルトグラム上の頂点座標の近似値 x'_i, y'_i 近傍で線形化すると、カルトグラム上の頂点座標を未知変数とする線形最小二乗問題で表される。この式(2)の繰り返し計算によって、与えられたデータをカルトグラム上で表現する頂点座標、すなわち地域形状を得ることができる。なお、表記上の都合より $x'_{ij} = x'_j - x'_i$, $y'_{ij} = y'_j - y'_i$ 、三角形 t_{ijk} の符号付き面積の近似値を $A'_{ijk} = (x'_{ij} y'_{ik} - x'_{ik} y'_{ij}) / 2$ と記す。

$$\begin{aligned} \min & \left[\sum_{t_{ijk} \in T} \left\{ 1 + \frac{3A'_{ijk}}{D_{ijk}^2} + \frac{A'_{ijk}}{D_{ijk}^2} (y'_{jk} x_i - y'_{ik} x_j + y'_{ij} x_k - x'_{jk} y_i + x'_{ik} y_j - x'_{ij} y_k) \right\}^2 \right. \\ & \left. + \mu \sum_{e_{mn} \in E} \left\{ \frac{(x_n - x_m) y'_{mn} - x'_{mn} (y_n - y_m)}{d'_{mn}^2} \right\}^2 \right] \end{aligned} \quad (2)$$

ここで、式(2)の解法に関する問題点について記す。式(2)の線形最小二乗問題は、観測方程式数 m は(三角形数)+(辺数)、未知変数数 n は(頂点数-1)×2 となる。QR 分解で線形最小二乗問題を解く場合、計算量は $O(mn^2)$ であるため、地域数の増加や地域形状の詳細化による頂点数の増加によって、計算時間が爆発的に増加する。また、メモリ使用量も急激に増加するため計算が困難で、大規模問題に対しては実用性が乏しい。

本研究では、計算量増大の問題の解決策として 2 つのアプローチを提案する。まず、カルトグラム作成計算に入力する地域数・頂点数の削減を、次に、式(2)が

疎行列の最小二乗問題であることを活かしたカルトグラム作成計算の実行を提案する。

3. 地域・頂点数を削減した作成手法

3.1 提案手法の全体構成

本章では、カルトグラム作成計算に入力する地域数・頂点数を減少させ、計算量を削減する手法を提案する。提案手法の全体構成は下記の通りである。

まず、隣接地域の集約化・地域形状の簡略化によって地域・頂点数の削減を行い、カルトグラム作成計算で地域形状変形を行う。繰り返し計算にて集約地域の面積をデータに合わせながら、削減された頂点の位置を空間内挿して全地域の形状を復元する。次に、集約地域毎に外周の頂点を固定した状態でカルトグラム作成計算を行い、集約地域内の各地域の面積をデータに合わせたカルトグラム作成を実行する。

提案手法により、一度に計算対象となる地域数・頂点数が減少するため、計算量を大幅な削減が可能である。個々の詳細について下記に記す。

3.2 提案手法の個別機能

3.2.1 隣接地域の集約・地域形状の簡略化

隣接地域の集約により地域数を減らすと同時に、集約地域の外周上以外の頂点を除外することができる。事前実験の結果、形状変形傾向(拡大率=データ/面積)の差が大きい地域を集約すると、無用な形状変形が生じたため、隣接地域の拡大率の差に閾値を設定し、条件を満たす隣接地域を集約する。

また、地域形状の簡略化で頂点数削減ができる。データに面積を合わせるカルトグラム作成問題の特性上、頂点の省略で地域面積が大きく変わることは望ましくないため、簡略化前後の面積比に閾値を設定し、条件を満たす最小頂点数を目指し頂点を省略する。具体的には、3 地域以上によって共有されている頂点だけの状態から、辺からの最遠点を追加する手順で面積比の条件を満たすまで点を追加する。

また、地域形状復元は後述の手順で行うが、省略される頂点は簡略化後地域のいずれかの内側に含まれていなければならない。そこで、全地域の外周上の頂点が簡略化後地域の外周辺上または内部に含まれているよう地域形状簡略化を行う。

3.2.2 地域形状の復元

カルトグラム作成計算後には、地域形状復元のため、隣接地域集約や地域形状簡略化により省略されカルトグラム上座標が求まっていない頂点座標を計算する必要がある。そこで、カルトグラム作成計算に使用した三角網を用いて三角形毎にアフィン変換を行って頂点座標を求める。但し、復元後に地域辺が交差する可能性がある(図1)ため、交差する場合には、変形前の地域辺と三角網辺の交点に新たな頂点を追加し、その交点で辺を曲げて地域辺の交差を回避する。

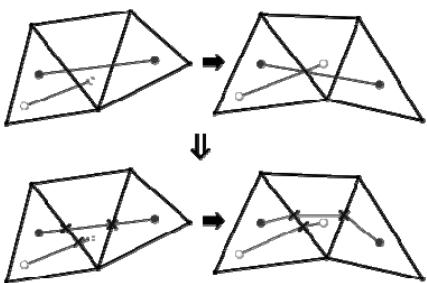


図1 頂点復元時の辺の交差に対する対応

3.2.3 周囲の頂点を固定したカルトグラム作成

集約地域の面積をデータに一致させた後、集約地域内の個々の地域形状をデータに合わせる必要がある。そこで、集約地域の外周の頂点座標を固定し、内部の頂点座標を変数として式(2)を解く。集約地域毎に別々に計算する事が可能なため、問題の規模を小さくする事が可能である。

3.3 提案手法の適用

提案手法を関東地方の2005年市区町村人口カルトグラムの作成に適用した(図2)。隣接関係にない地域の重なりを回避するため、また島嶼部を他の地域と接続するため、東京湾など一部の水域にダミーデータを設定している。地域数は319、頂点数は2638である。

まず、隣接地域の中から、拡大率(データ/面積)の差が0.1以下の地域を集約する(図3)。この過程で、地域数は274、頂点数は2442に削減された。

次に、地域形状を簡略化する(図4)。簡略化前後で面積の違いが10%以下となるよう頂点を省略した結果、頂点数を962まで削減できた。

その後、図4の地域形状を入力しカルトグラム作成計算を実行した。カルトグラム作成計算では、三角網の反転が生じると三角網を組み直す必要があるが、そ

の度に地域形状復元・隣接地域集約・地域形状簡略化を繰り返し実行した。なお、繰り返し過程で、簡略化時の面積に関する閾値の値を小さくし、より正確に集約地域の面積をデータに合わせる処置を行っている。最終的に、集約地域毎に外周上頂点を固定してカルトグラム作成計算を行い、図5(a)のカルトグラムを得た。また、図5(b)に既存手法による結果を示す。

なお、計算時間は、CPU Xeon2.66GHz×2の計算機で提案手法は196秒、既存手法は521秒である。また、最大使用メモリは、提案手法は193Kbyte、既存手法は606Kbyteである。このように、計算時間・使用メモリ領域共に改善が見られた。しかし、全国の市区町村人口カルトグラム作成には、約1時間半を要し、実用性には限界が見られた。



図2 関東地方市区町村形状データ

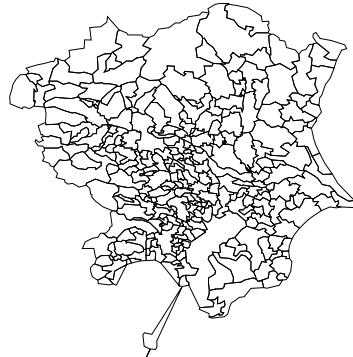


図3 隣接地域の集約(拡大率の差 0.1 以下を集約)



図4 地域形状の簡略化(面積の違い 10%以下)

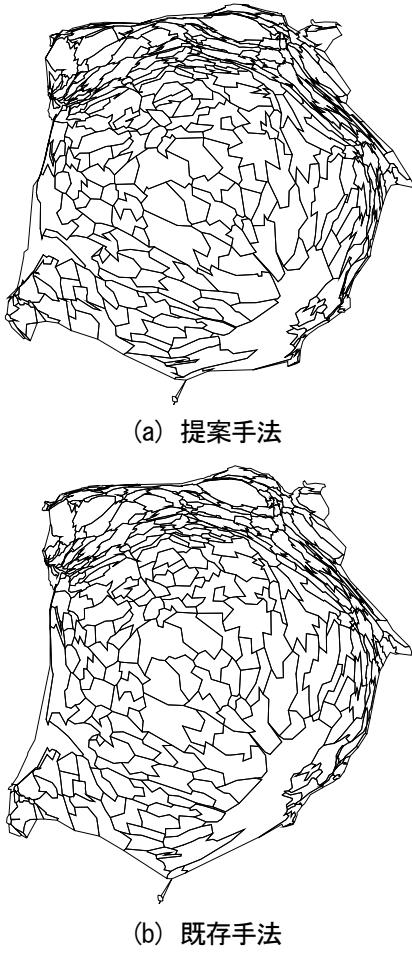


図 5 提案手法と既存手法による作図結果

4. 疎行列を利用する解法

式(2)は、疎行列で定義される最小二乗問題である。そこで、Intel Math Kernel Library の直接法スパースソルバーPARDISO を利用し、疎行列を活かした数値計算を実行し、短時間のカルトグラム作成を目指す。

まず、PARDISO を利用した計算を行うためには、式(2)の最小二乗問題を連立方程式問題に変更する必要がある(Intel, 2006)。そこで、式(2)を変数(カルトグラム上の頂点座標 x_i, y_i)で偏微分した値が 0 となる x_i, y_i を求める問題として書き換えた。この式展開によって、観測方程式数・未知変数数が共に(頂点数-1)×2 となる連立方程式を解く問題と変更された。

3.3 で使用したものと同じ計算機・データを用いて適用したところ、約 6 秒でカルトグラム作成が終了した。連立方程式の行列のうち、非 0 要素は 0.25% にすぎず、疎行列という特徴を活かした解法が非常に効果的であることが分かる。

また、より大規模な問題である日本全国の市区町村人口データに対して適用を行った。本州・北海道・四国・九州を別々に計算しているが、最もデータが大規模な本州(地域数 1453、頂点数 11066)の場合でも、294 秒の計算時間で出力することができた(図 6)。

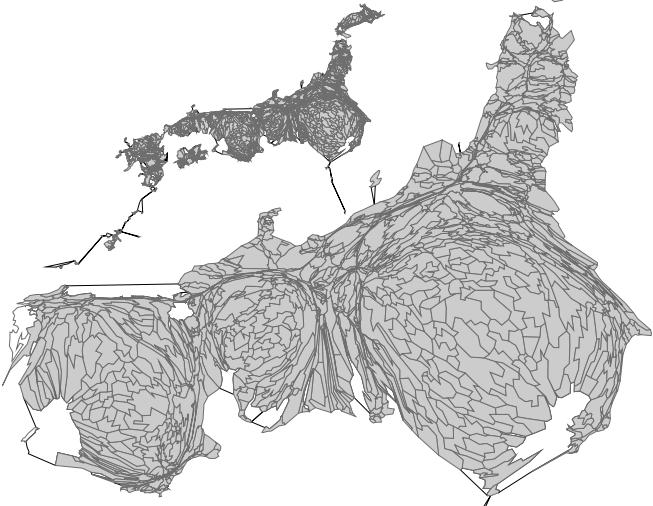


図 6 既存手法の数値計算法改良による
日本の市区町村人口カルトグラム(一部抜粋)

5. おわりに

本研究では、2 つのアプローチでカルトグラム作成問題の計算量を削減する試みを行った。双方とも計算時間の短縮に効果的であったが、特に疎行列に対応した数値計算手法の利用が非常に効果的であった。

なお、提案の 2 手法を組み合わせた作成手法も構築したが、前者の隣接地域集約・地域形状簡略化・地域形状復元に時間を要するため、疎行列の数値演算法を活用した後者の手法単独によるカルトグラム作成の計算時間を超えることができなかった。数値計算を利用する他の研究にこの結果が参考になれば幸いである。

なお、疎行列の数値演算法を活用した既存のカルトグラム作成手法のソースコードを下記のアドレスにて公開する。

<http://planner.t.u-tokyo.ac.jp/member/inoue/carotram/>

参考文献

- 井上 亮, 清水英範 (2005) 連続エリアカルトグラム作成の新手法—GIS時代の統計データの視覚化手法—, 土木学会論文集, No.779(IV-66), pp.147-156.

Intel (2006) Intel Math Kernel Library Reference Manual, Intel.