

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 -

井上 亮¹・清水英範²

¹学生会員 修(工) 東京大学大学院博士課程 工学系研究科社会基盤学専攻
(〒113-8656 東京都文京区本郷7-3-1)

E-mail: inoue@trip.t.u-tokyo.ac.jp

²正会員 工博 東京大学大学院教授 工学系研究科社会基盤学専攻
(〒113-8656 東京都文京区本郷7-3-1)

E-mail: shimizu@civil.t.u-tokyo.ac.jp

連続エリアカルトグラムとは、地図上の地域の面積が地域の統計データの大小を表すように地理的地図を変形した地図である。近年、GIS時代の統計データの視覚化手法として注目されている。

連続エリアカルトグラム作成問題は、地域の面積に関する情報だけから形状を定めることはできないため、解の一意性を持たない不良設定問題である。その解法は様々提案されてきたが、数学的に明快で、操作性に優れ、短時間で計算可能な、視覚化手法として有効な解法は未だ提案されていない。

本論文では、以上の課題を解決する数学的に明快で簡潔かつ高速な連続エリアカルトグラム作成アルゴリズムを提案し、実用性の高い連続エリアカルトグラム作成手法を構築した。

Key Words: *continuous area cartogram, visualization, triangulated irregular network*

1. はじめに

近年、国や地方自治体等により収集された多種多様な統計データが電子媒体として無料あるいは安価に提供されるようになってきている。その結果、一般市民や分析者が多くのデータを利活用できる環境が整ってきた。これらの統計データ群は地理情報に基づいてGIS上で管理される。このGISの基本機能の一つとして統計データの視覚化機能が提供されており、プレゼンテーションやデータマイニングの道具として利用されている。従来、GISソフトウェアには、地図を統計データの値に基づき色分けするコロプレスマップや、地図上に点を表示し分布を表現するドットマップ、棒グラフや円グラフ等を地図上に記載する手法等が実装されてきた¹⁾。

統計データの視覚化手法にはこれらの他に、計量地理学やコンピュータ科学で注目されているカルトグラムと呼ばれる手法がある²⁾。カルトグラムとは、各地域の統計データに基づき地図を変形し、地域の特徴を表現する主題図である。このカルトグラムの一手法に、各地域の

統計データの大小をカルトグラム上の面積の大小を用いて表現するように通常の地理的地図を変形させて表示する、連続エリアカルトグラムと呼ばれる手法がある³⁾。連続エリアカルトグラムは、エリアカルトグラム、あるいは単にカルトグラムと呼ばれることも多い。本論文では以後、連続エリアカルトグラムをエリアカルトグラムと表記する。

エリアカルトグラムは統計データの地理的分布を印象的に視覚化することができ、読図者が地域の特徴を直感的に理解することを促す。GIS環境へ実装することにより、カルトグラム作成手法は、GIS上で管理されている大量の統計データを用いたデータマイニングに威力を発揮し、統計データのより有効な利活用を促す道具となることが期待される。

エリアカルトグラムには今まで数多くの作成手法⁴⁾⁻⁹⁾が提案されてきた。しかし、複雑なパラメータ設定が必要なものや、計算に時間を要するもの、データを高精度で表現できないものなど様々な問題を抱えている。また、これらの作成手法が一般のユーザーが利用できるように

GISの拡張機能等のソフトウェアとして提供されている例¹⁰⁾はほとんど見られないのが現状である。

エリアカルトグラム作成問題とは、面積が与えられた値に一致するように地域形状を変形する問題である。この実行には下記の3つの課題がある。

エリアカルトグラム上の地域の面積に関する情報だけでは地域形状を一意に定めることは不可能である。つまり、エリアカルトグラム作成問題は解の一意性を持たない不良設定問題であり、何らかの正則化条件を設定する必要がある。

一意な解を求めるだけなら適当な正則化条件の導入で解決できる。しかし、エリアカルトグラムを視覚化手法として利用する際には、地理的地図との対比を通して地域形状の違いを観察し、エリアカルトグラム上に表現された統計データを解釈することになる。その時、例えば地域の位置関係の反転等の、統計データの表現には不必要な変形が含まれていると、エリアカルトグラム上に表現された統計データの解釈を妨げることになる。そのため、エリアカルトグラム作成に当たっては、統計データを面積で表現するように地域形状を変形すると同時に、地理的地図からの地域形状の変形を抑える必要がある。

エリアカルトグラム作成手法をソフトウェアとして社会一般に提供する場合には、パラメータ設定等に多くの試行錯誤をソフトウェア利用者に強いる手法や長時間の計算を要する手法は実用的ではない。特に、複雑なアルゴリズムを用いた作成手法では、利用者にとってはブラックボックスとなってしまう、パラメータの意味解釈ができないため合理的なパラメータ設定が行えない。このような作成手法は操作性に乏しいと言わざるを得ない。真にユーザーフレンドリーな手法とは、作成アルゴリズムが数学的に明快でパラメータの意味解釈が容易な手法であると筆者らは考える。

本論文では、以上の課題を解決するエリアカルトグラム作成手法を提案する。具体的には、簡潔な手法の構築を行うため、任意形状の地域を三角網分割することを前提とし、三角形の面積を統計データに合わせるよう変形するアルゴリズムを提案する。また、正則化項として地理的地図からの変形を抑制する条件を導入し、視覚化された統計データの解釈が容易なエリアカルトグラム作成を行うことを目指す。

2. 既存のエリアカルトグラム作成手法

本章では、エリアカルトグラム作成手法に関する既存研究について記し、その問題点を述べる。

Tobler⁴⁾では、エリアカルトグラム上の座標を (x, y) 、通常の地図上の座標を (u, v) 、エリアカルトグラムで表現する統計データの (u, v) 上での密度を $h(u, v)$ とおき、変形の少ない一意な解を求めるため、角度変化を最小化するDirichlet積分最小化を設定し、エリアカルトグラム作成問題を式(1)と定義している。

$$\min \int_R \left(\frac{\partial x^2}{\partial u} + \frac{\partial y^2}{\partial u} + \frac{\partial x^2}{\partial v} + \frac{\partial y^2}{\partial v} \right) dudv \quad (1)$$
$$s.t. \quad \frac{\partial x}{\partial u} \frac{\partial u}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial u}{\partial u} = h(u, v)$$

実際にエリアカルトグラム作成するには、地域を正方形格子に分割した後、式(1)の近似解法を各格子に適用し頂点の移動量を計算している。しかし、この近似解法では統計データの表現に不必要な地域形状の変形を抑えることができず、頂点の位置関係が反転し位相が破壊されることを防げない。そこで、頂点の座標変化量を1点ずつ個別に計算し、位相破壊が生じる場合には座標変化量を小さくするという場当たり的な手法で調整している。また、簡便解法⁵⁾として xy 方向に独立に計算する手法も提案しているが、統計データを高精度に表現できない。

Dougenik *et al.*⁶⁾ではTobler⁴⁾の解法を改良し高速化を果たしているが、アルゴリズムは不明快さを拭えず、また必ずしも視覚的に分かりやすいエリアカルトグラムを作成できない。また、Gusein-Zade・Tikunov⁷⁾では、各地域の重心・頂点間距離を統計データに基づき伸縮させる手法を提案している。この手法では地域形状が重心を中心とする円に近づくため形状変化が大きく、視覚的に優れたエリアカルトグラムを作成できない。

これまでの手法では、視覚的に分かりやすいエリアカルトグラムを作成するために統計データの表現に不必要な地域形状変化を抑えることを十分に考慮していなかった。しかし、Kocmoud・House⁸⁾では、統計データに面積を合わせる計算と地域形状変化を抑える計算を交互に行い、視覚的に分かりやすいエリアカルトグラム作成を行う手法を提案している。この手法では、地域形状を可能な限り保持しながら高い精度でデータを表現するエリアカルトグラムを作成できる。しかし、そのアルゴリズムは、地域形状維持のために「辺の方位角変化抑制」「辺長の比率変化抑制」を、位相破壊の防止のために「地域形状の内角変化抑制」「辺長変化抑制」「辺の交差防止」を導入するなど、都合5つの制約条件を設定してエリアカルトグラム作成を行っている。これらの制約条件は複雑で数学的に明快ではなく、またそれらの重み付けも論文中で明らかにされていない。また、本手法によるとアメリカ合衆国州人口エリアカルトグラムの作成にCPU300MHzの計算機で6時間も計算を要し、視覚化手法として実用的とは言えない。

Keim *et al.*⁹⁾では、CartoDrawと呼ぶ作成アルゴリズムを提案している。Kocmoud・House⁸⁾と同様に、統計データの面積による表示とともに地域形状変化の抑制を考慮したアルゴリズムを構築しており、視覚的に分かりやすいエリアカルトグラムを作成することが可能である。さらに、Kocmoud・House⁸⁾より約2000倍高速に計算できるとしており、既存の最良手法といえる。

CartoDrawのアルゴリズムでは、scanlineと呼ぶ直線を設定し、地域の面積を統計データに合わせるようにscanlineから一定距離内の頂点をscanlineの垂線方向に移動させる手法を提案している。但し、地域形状変化が基準を超えた場合は変形を止めるというルールを設定している。このアルゴリズムの難点は、初期設定であるscanlineに対する結果の依存性が強くこの設定によりエリアカルトグラム形状が大きく異なる点、また不適切に設定された場合には統計データを低い精度で表現するエリアカルトグラムしか作成できない可能性がある点である。

このように、既存のエリアカルトグラム作成手法は、視覚的に分かりやすいエリアカルトグラムを作成できるようになってきているが、アルゴリズムが複雑な手法や、初期値設定に多くの工夫を要する手法等、未だ実用的とは言えない。

3. エリアカルトグラム作成手法の提案

本章では、エリアカルトグラム作成手法の提案に向けて、作成手法の基本的な考え方をまとめた後、新たなエリアカルトグラム作成手法の提案を行う。

(1) 基本的な考え方

エリアカルトグラム作成上の問題点を整理し、その解決策として、三角網分割を用いて三角網上の辺の方位角変化に関する正則化条件を設定することを提案する。

エリアカルトグラム作成上の問題点は、「適用対象の地域形状が多様」、「エリアカルトグラム作成問題は不良設定問題」、「地域形状変形の抑制が必要」の3つが挙げられる。

そこで本研究では、「適用対象地域の三角網分割」と「三角網の辺の方位角変化を抑制する正則化項の導入」によって上記の問題点に対応する。

まず、「対象地域の三角網分割」によって、問題点「適用対象の地域形状が多様」を解消する。もともと適用対象の地域形状は多種多様であるが、複雑な形状の面積計算は容易ではなく、あらゆる形状に適用可能な手法の構築は容易ではない。しかし、適用対象地域を三角網分割することにより、三角形をデータに合わせて変形する手法を構築すればよい。三角網分割を用いる利点は、

三角形が最も単純な形状で面積計算が容易な点である。このため、面積をデータに合わせる目的関数を簡潔に記述できる。

次に、「正則化項の導入」により、問題点「エリアカルトグラム作成問題は不良設定問題」を解決する。面積に関する情報だけでは地域形状を定めることは不可能であるから、エリアカルトグラム作成問題には正則化条件が必要である。また、「三角網の辺の方位角変化を抑制する正則化項の導入」により問題点「地域形状変形の抑制が必要」に対応し、表現する統計データに因らない不必要な形状の変形を排除し、地理的地図との比較が容易な視覚的に優れたエリアカルトグラムを作成する。この正則化条件により三角網の変形、つまり地域形状の変形を抑えることが可能になる。例えば、方位角変化の抑制に対して極端に大きな重み付けを行った場合、通常の地理的地図から拡大・縮小のみを許容することになり、形状を完全に維持した変形を行うことができる。

以上から、本論文では地域を三角網分割し、三角網を構成する辺の方位角変化に対する正則化項を導入し、三角形の面積を与えられたデータに合わせる手法の構築を行う。

(2) 三角網分割を用いた作成手法の提案

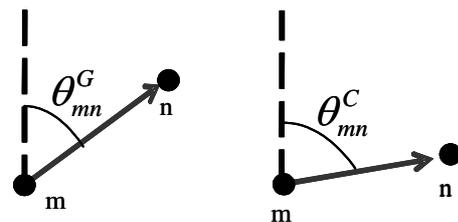
頂点 i, j, k からなる三角形 t_{ijk} のエリアカルトグラム上の面積 A_{ijk} をデータ D_{ijk} に合わせる目的関数は、 T を三角網に含まれる三角形の集合とすると、式(2)と書ける。

$$\min \sum_{t_{ijk} \in T} (D_{ijk} - A_{ijk})^2 \quad (2)$$

もちろん式(2)では解が無数に存在するため、正則化項を導入する必要がある。正則化項として、地域形状変化を抑えるため、三角網の辺の方位角を拘束することとする。辺 mn の地理的配置上の座北方位角を θ_{mn}^G 、エリアカルトグラム上の座北方位角を θ_{mn}^C (図-1)、頂点 mn を結ぶ辺を e_{mn} 、三角網に含まれる辺の集合を E とすると、辺 mn の方向変化を抑制する正則化項は式(3)となる。

$$+ \sum_{e_{mn} \in E} (\theta_{mn}^C - \theta_{mn}^G)^2 \quad (3)$$

主問題(式(2))と正則化項(式(3))より、三角網を用いたエリアカルトグラム作成問題は、正則化項の重みを μ



(a) 地理的配置上 (b) エリアカルトグラム上
図-1 辺 mn の座北方位角

とすると、式(4)となる。

$$\min \left[\sum_{t_{ijk} \in T} (D_{ijk} - A_{ijk})^2 + \mu \sum_{e_{mn} \in E} (\theta_{mn}^C - \theta_{mn}^G)^2 \right] \quad (4)$$

しかし式(4)では、主問題と正則化項の次元が異なり μ の設定が難しい。また、より大きなデータが与えられた三角形の面積をより正確に合わせることを重視した式となっている。エリアカルトグラム作成では、データの大小に関わらずデータを面積で正確に表現するように三角形の変形を行うことが望ましい。そこで、式(5)のようにエリアカルトグラム作成の主問題を無次元化する。

$$\min \left[\sum_{t_{ijk} \in T} \left(1 - \frac{A_{ijk}}{D_{ijk}} \right)^2 + \mu \sum_{e_{mn} \in E} (\theta_{mn}^C - \theta_{mn}^G)^2 \right] \quad (5)$$

ここで、頂点 m のエリアカルトグラム上の座標を (x_m, y_m) 、地理的地点配置上の座標を (x_m^G, y_m^G) と表すこととし、式(5)を三角網の頂点座標を用いた式に改める。

まず、エリアカルトグラム上の三角形の面積 A_{ijk} はその頂点 i, j, k の座標を用いると、式(6)と表せる。

$$A_{ijk} = \frac{1}{2} |(x_j - x_i)(y_k - y_i) - (x_k - x_i)(y_j - y_i)| \quad (6)$$

主問題(式(5) 第一項)に式(6)を代入する際に、絶対値記号を取り除くため、主問題を式(7)と改める。

$$\min \sum_{t_{ijk} \in T} \left(1 - \frac{A_{ijk}^2}{D_{ijk}^2} \right)^2 \quad (7)$$

式(7)は式(5)第一項と比べ面積 A_{ijk} のデータ D_{ijk} からのずれをより大きく評価するが、エリアカルトグラム作成の主問題として不適切ではない。式(7)に式(6)を代入する。

$$\min \sum_{t_{ijk} \in T} \left[1 - \frac{1}{4D_{ijk}^2} \{(x_j - x_i)(y_k - y_i) - (x_k - x_i)(y_j - y_i)\}^2 \right] \quad (8)$$

また、正則化項(式(5) 第二項)は、

$$+ \mu \sum_{e_{mn} \in E} \left\{ \arctan \left(\frac{x_n - x_m}{y_n - y_m} \right) - \arctan \left(\frac{x_n^G - x_m^G}{y_n^G - y_m^G} \right) \right\}^2 \quad (9)$$

と書けることから、

$$+ \mu \sum_{e_{mn} \in E} \left\{ \frac{x_n - x_m}{y_n - y_m} - \frac{x_n^G - x_m^G}{y_n^G - y_m^G} \right\}^2 \quad (10)$$

とほぼ同値となる。更に、計算上分母が 0 となるのを避けるため、式(11)とする。但し、頂点 mn 間距離を d_{mn} と記述する。

$$+ \mu \sum_{e_{mn} \in E} \left\{ \frac{(x_n - x_m)(y_n^G - y_m^G) - (x_n^G - x_m^G)(y_n - y_m)}{d_{mn} d_{mn}^G} \right\}^2 \quad (11)$$

以上より、エリアカルトグラム作成問題は式(12)と書ける。

$$\min \left[\sum_{t_{ijk} \in T} \left\{ 1 - \frac{\{(x_j - x_i)(y_k - y_i) - (x_k - x_i)(y_j - y_i)\}^2}{4D_{ijk}^2} \right\}^2 + \mu \sum_{e_{mn} \in E} \left\{ \frac{(x_n - x_m)(y_n^G - y_m^G) - (x_n^G - x_m^G)(y_n - y_m)}{d_{mn} d_{mn}^G} \right\}^2 \right] \quad (12)$$

式(12)の非線形最小二乗問題を解くことにより、エリアカルトグラムを作成することができる。但し、位相関係を保ったエリアカルトグラム作成のためには、正則化項の重み μ を適切に設定し、三角網の反転を防ぐ必要がある。

ここで、式(12)で一意に解を求められることを確認する。未知変数の数は、1 点を座標系に固定するため、(頂点数-1) \times 2 となる。一方、観測方程式数は、(三角形数) + (辺数)となる。三角網では常に、(未知変数数) (観測方程式数) が成り立つので、式(12)から解を一意に求めることができる。

(3) 解法の線形化

エリアカルトグラム上の座標 x_i, y_i に近似値 x'_i, y'_i を与え、式(12)を近似値近傍で線形化する。なお、表記上の都合より、以後

$$x'_{ij} = x'_j - x'_i, \quad y'_{ij} = y'_j - y'_i \quad (13)$$

と表し、符号付きの三角形 t_{ijk} の面積の近似値を

$$A'_{ijk} = \frac{1}{2} (x'_{ij} y'_{ik} - x'_{ik} y'_{ij}) \quad (14)$$

と表記する。

式(12)の主問題を、近似値近傍で線形化すると

$$\sum_{t_{ijk} \in T} \left\{ 1 + \frac{3A'_{ijk}{}^2}{D_{ijk}^2} + \frac{A'_{ijk}}{D_{ijk}^2} (y'_{jk} x'_i - y'_{ik} x'_j + y'_{ij} x'_k - x'_{jk} y'_i + x'_{ik} y'_j - x'_{ij} y'_k) \right\}^2 \quad (15)$$

となる。また、式(12)の正則化項は、繰り返し計算時に近似値を用いて正則化条件を更新すると、

$$+ \mu \sum_{e_{mn} \in E} \left\{ \frac{(x_n - x_m) y'_{mn} - x'_{mn} (y_n - y_m)}{d'_{mn}{}^2} \right\}^2 \quad (16)$$

と表せる。式(15)(16)より、式(12)は

$$\min \left[\sum_{t_{ijk} \in T} \left\{ 1 + \frac{3A'_{ijk}{}^2}{D_{ijk}^2} + \frac{A'_{ijk}}{D_{ijk}^2} (y'_{jk} x'_i - y'_{ik} x'_j + y'_{ij} x'_k - x'_{jk} y'_i + x'_{ik} y'_j - x'_{ij} y'_k) \right\}^2 + \mu \sum_{e_{mn} \in E} \left\{ \frac{(x_n - x_m) y'_{mn} - x'_{mn} (y_n - y_m)}{d'_{mn}{}^2} \right\}^2 \right] \quad (17)$$

のように線形化できる．式(17)の線形最小二乗問題の繰り返し計算によりエリアカルトグラム作成ができる．

(3) 提案解法に基づくエリアカルトグラム作成手法

式(17)の解法を用いてエリアカルトグラム作成を行う際には，次の2点を考慮する必要がある．

- ・ 全ての三角形の反転を防ぐ．
- ・ 実用的な計算時間で図を出力するため，解法を高速化する．

これら2点は，正則化項に対する重み μ の大きさを制御することで，解決することができる．その詳細を以下に示す．

まず， μ に小さな値を設定した場合，三角網の辺の方向変化が大きくなるため，三角形の反転が生じる可能性が高くなる．三角形の反転が生じると，図-2の左図から右図の変形のように頂点の位相が破壊される．三角形の反転を防ぐためには， μ の値を大きく設定し三角網を構成する各辺の方向変化に対する制約を大きくすることが望ましい．

一方， μ に大きな値を設定すると，収束速度の低下をもたらす，収束計算回数・計算時間の増加を招く．また，収束計算過程において μ を固定していると，式(17)の主問題(第一項)の値は収束計算毎に小さくなるため，相対的に正則化条件(第二項)が大きくなり，収束速度が低下する．式(17)は，1次収束する手法であるため，最適解近傍では収束速度が低下し，収束計算に時間を要する．この方法を超1次収束させるためには，最適解近傍で μ 0となるよう制御する必要がある¹¹⁾．

三角形の反転を防ぎ，かつ，計算時間の短いエリアカルトグラム作成を行うためには，収束計算1回毎に，三角形の反転が起こらない最小の μ 値を設定することが望ましい．しかし，この値を求めることは困難である．そこで， μ の値を

$$\left\{ \sum_{i_{jk} \in T} (D_{ijk} - A_{ijk})^2 / \sum_{i_{jk} \in T} (D_{ijk} - A'_{ijk})^2 \right\} \mu \quad (18)$$

に基づき減少させながら収束計算を行い，三角形の反転が観測された場合には μ を増加させて再計算を行い，三角形反転の防止，および，計算時間の短縮を図る．

なお，三角形反転は，頂角の正弦の符号変化により判断することができる．

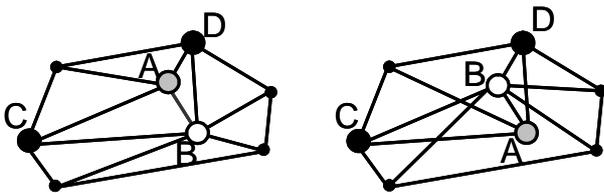


図-2 ABC・ABDの反転と三角網の位相破壊

(4) 三角網の自動作成

提案アルゴリズムを用いてエリアカルトグラム作成を行うためには，三角網を作成する必要がある．そのため，三角網の自動作成手法を用いる．

エリアカルトグラム作成上望ましい三角網とは，三角形の反転が起こりにくいものである．そのため，できる限り極端な鋭角を持たない三角形を作成することが望ましい．そこで，全ての三角形分割の中で，三角網の最小角が最大となるDelaunay三角分割¹²⁾を用いる．

三角網自動作成の実装に当たっては，まず各地域のポリゴンをその頂点を用いてDelaunay三角網を作成し，その後，内点の追加・ラブラシアン法による内点座標の修正を用いて三角網を作成する¹³⁾．

また，エリアカルトグラム作成の高速化のため，提案アルゴリズム内で一定回以上三角形の反転が生じた場合，三角網を修正することとする．

なお，本提案手法によるエリアカルトグラム作成アルゴリズムの手順は図-3の通りである．

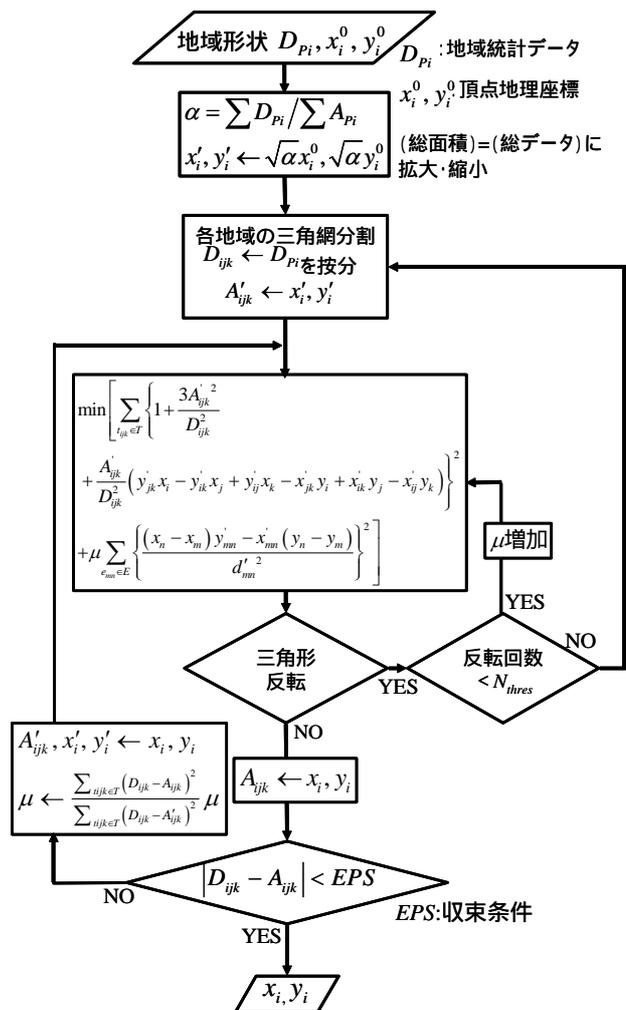


図-3 連続エリアカルトグラム作成アルゴリズムのフローチャート

4. 提案手法の適用

本提案手法を実際のデータに適用し、その適用可能性を検証する。

(1) 適用と評価 -アメリカ合衆国人口データ-

U.S. Census Bureau, United States Department of Commerce による 1980 年アメリカ合衆国州人口データを用いてエリアカルトグラムを作成する。本提案手法適用に際して、簡略化したアメリカ合衆国地図(図-4)を用いる。頂点数は 185 である。但し、本提案手法では全地域は連続していないと計算できないため、島がある場合には海峡に仮想的なポリゴンを作り仮想的な値を与えている。また、湾を挟む両岸の地域は交差する可能性があるため、湾にも仮想的なポリゴンを設定している。また、地域が複数のポリゴンに分かれている場合、人口データを地理的な面積に基づいて按分している。なお、エリアカルトグラム作成計算では、1 点をエリアカルトグラム座標に固定する必要がある。本手法では任意の点を固定しても出力されるエリアカルトグラム形状は変わらないので、 x 軸の最小値をとる点を固定して計算した。

正則化項に対する重み μ の初期値を 1, $\sqrt{10}$, 10 と設定しエリアカルトグラム作成計算を行った。その結果を図-5, 収束計算回数・計算時間・RMSE を表-1 に示す。また、人口データと各エリアカルトグラム上に表現された面積の比較を表-2 に示す。また、 μ の初期値 $\sqrt{10}$ の場合の計算過程を図-6 に示す。

まず、本提案手法により、エリアカルトグラム作成が実用的な計算時間で計算可能であることが確認できた。1 州あたりの人口が約 460 万人であることを考えると RMSE は十分に小さく、本提案手法によって得られた図は人口データを表現していることが分かる。なお、表-2 のとおり、人口データとエリアカルトグラム上に表現された面積の間の相関係数はいずれも 0.999 を超えている。

次に、 μ 初期値の与え方による結果の違いについて注目する。表-1 より、 μ 初期値が小さいと収束速度が速く、計算時間が短いことがわかる。しかし、図-5(a)より、 μ 初期値に 1 を与えた場合には、非常に変形が大きく分かりにくいエリアカルトグラムとなってしまう。一方、図-5(b)(c)を見ると、地域形状は比較的保たれており、分かりやすい。このように、 μ の初期値を大きくすることにより辺の角度変化を抑えることが可能になり、比較的分かりやすいエリアカルトグラムを作成することが確認できた。また図-5(b)(c)間には目立った差異は見られないが、計算時間では約 3 倍の差がある。このことから、正則化項に対する過剰な重み付けは出力結果にはあまり影響を与えず、むしろ計算時間の増加をもたらす好ましくない。しかし、形状変化を抑えつつ高速に計算できる

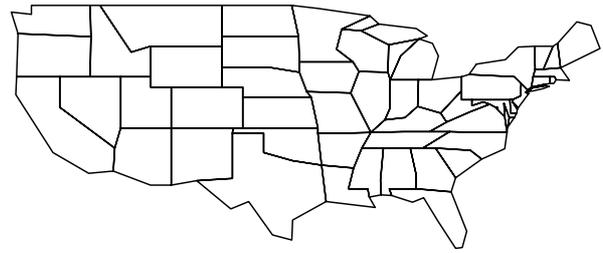
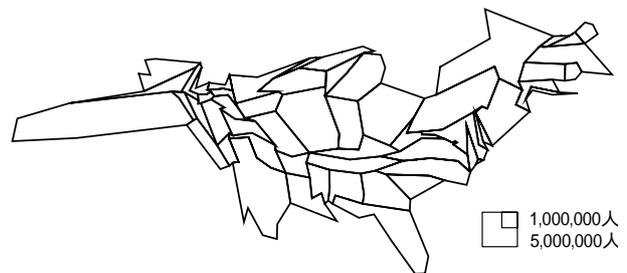


図-4 入力したアメリカ合衆国 州形状データ

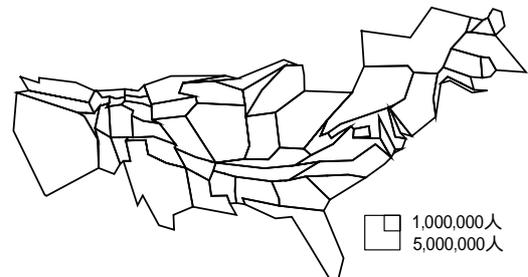
表-1 1980年アメリカ人口エリアカルトグラム作成
収束計算回数・計算時間・RMSE

μ 初期値	収束計算回数 (回)	計算時間 (秒)	RMSE (人)
1	9	3.5	8.8×10^4
$\sqrt{10}$	23	9.2	5.1×10^3
10	100	29.7	2.7×10^3

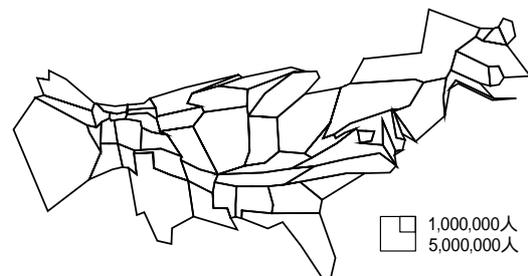
(但し、CPU1.8GHz 使用)



(a) μ 初期値 1



(b) μ 初期値 $\sqrt{10}$

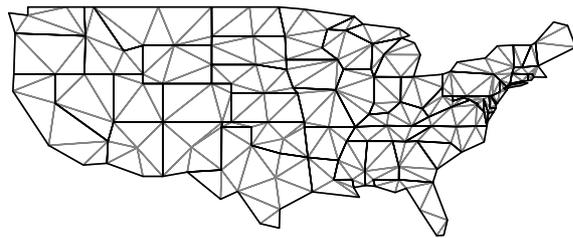


(c) μ 初期値 10

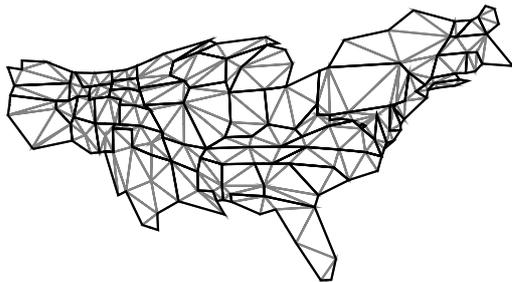
図-5 提案手法による 1980 年アメリカ合衆国
州人口エリアカルトグラム

μ の最適な値を求めることは現状では困難であり，試行錯誤に頼らざるを得ない。

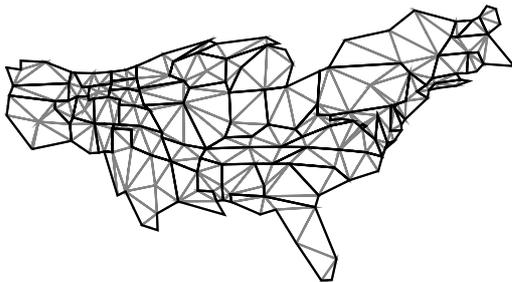
なお，計算過程は図-6 に示す通りである．まず各地域を三角形に分割を行う(図-6(a))．エリアカルトグラム作成の繰り返し計算 4 回目終了後(図-6(b))，三角形の反転が生じるようになったため三角網の修正(図-6(c))を行っている．この三角網を用いて繰り返し計算を 12 回行い図-6(d)を得て，再び三角網の修正を行った後 7 回の繰り返し計算し，図-5(b)のエリアカルトグラムを出力する．



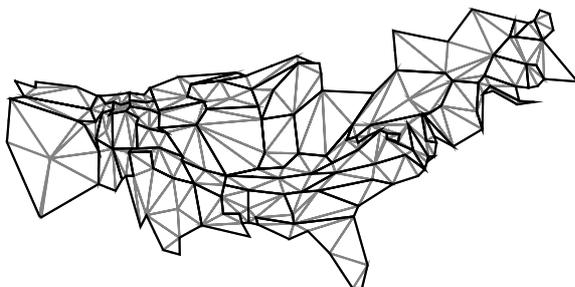
(a) 三角網の自動作成



(b) 繰り返し計算 4 回目終了後



(c) 三角網の修正



(d) 繰り返し計算 16 回目終了後

図-6 エリアカルトグラム 計算経過 ($\mu = \sqrt{10}$)

表-2 データの表現精度

州	人口	$\mu=1$	$\mu=\sqrt{10}$	$\mu=10$
AL	3,894,025	3,894,729	3,893,132	3,893,989
AZ	2,716,546	2,716,346	2,716,193	2,716,543
AR	2,286,357	2,286,216	2,286,718	2,286,532
CA	23,667,764	23,496,735	23,665,500	23,667,660
CO	2,889,735	2,888,109	2,889,254	2,889,670
CT	3,107,564	3,107,538	3,107,154	3,107,563
DE	594,338	594,265	592,248	594,095
DC	638,432	32,088	638,192	638,716
FL	9,746,961	9,746,115	9,735,179	9,746,443
GA	5,462,982	5,462,211	5,464,379	5,461,941
ID	944,127	934,257	944,023	944,126
IL	11,427,409	11,423,344	11,428,513	11,426,871
IN	5,490,210	5,489,232	5,490,021	5,488,999
IA	2,913,808	2,917,048	2,912,516	2,913,668
KS	2,364,236	2,364,681	2,363,710	2,364,189
KY	3,660,324	3,660,343	3,659,824	3,659,626
LA	4,206,116	4,204,324	4,206,016	4,205,999
ME	1,125,043	1,125,043	1,125,009	1,125,043
MD	4,216,933	4,215,346	4,216,516	4,233,617
MA	5,737,093	5,737,072	5,736,946	5,737,093
MI	9,262,044	9,258,050	9,259,392	9,259,061
MN	4,075,970	4,074,383	4,073,032	4,075,477
MS	2,520,770	2,520,158	2,520,800	2,520,784
MO	4,916,766	4,914,219	4,916,357	4,917,002
MT	786,690	784,501	786,541	786,664
NE	1,569,825	1,569,901	1,569,669	1,569,788
NV	800,508	834,208	800,417	800,505
NH	920,610	920,608	920,547	920,610
NJ	7,365,011	7,364,307	7,344,574	7,364,480
NM	1,303,302	1,299,711	1,303,101	1,303,290
NY	17,558,165	17,557,831	17,557,747	17,556,165
NC	5,880,095	5,880,344	5,873,931	5,882,543
ND	652,717	650,142	652,539	652,687
OH	10,797,603	10,792,102	10,791,224	10,793,150
OK	3,025,487	3,025,259	3,024,984	3,025,481
OR	2,633,156	2,608,879	2,632,955	2,633,154
PA	11,864,720	11,863,486	11,843,483	11,861,039
RI	947,154	947,148	947,112	947,154
SC	3,120,729	3,121,948	3,126,826	3,120,290
SD	690,768	690,440	690,661	690,738
TN	4,591,023	4,590,737	4,589,529	4,591,942
TX	14,225,513	14,223,969	14,223,606	14,225,449
UT	1,461,037	1,458,984	1,460,811	1,461,028
VT	511,456	511,444	511,426	511,456
VA	5,346,797	5,347,573	5,344,301	5,352,277
WA	4,132,353	4,111,747	4,132,124	4,132,352
WV	1,950,186	1,950,201	1,951,749	1,948,765
WI	4,705,642	4,711,027	4,706,041	4,705,363
WY	469,557	468,478	469,438	469,551
相関係数		0.9998	1.0000	1.0000

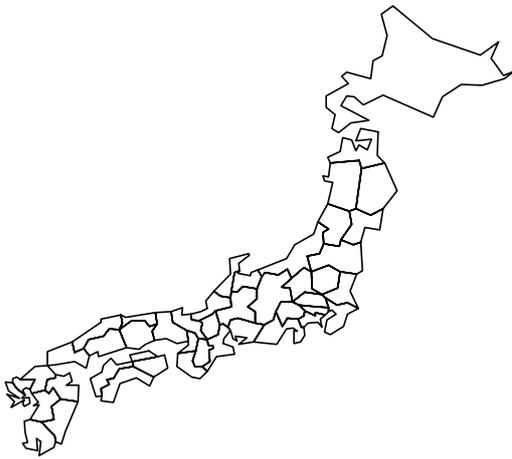


図-7 入力した日本 都道府県形状データ

(2) 視覚化例の提示 - 都道府県人口データ -

総務省統計局により整備された 1920~2000 年の 5 年毎の都道府県人口データに対して提案手法を適用し、人口の時系列変化を視覚化した例を示す。

都道府県人口の時系列変化を分かりやすく表現するためには、人口の増減に起因しない地域形状の変形を排除することが望ましい。そこで、まず 1920 年の人口エリアカルトグラムを、331 個の頂点からなる簡略化された都道府県形状データを用いて作成した(図-7)。その後、5 年毎の人口エリアカルトグラムを一時点前のエリアカルトグラムの地域形状を初期値として入力し作成した。

その中から 1940~2000 年の 10 年毎の都道府県人口エリアカルトグラムを示す(図-8)。但し、一時点前の人口データからの 1 年あたり人口増加率に応じて各都道府県を色分けしている。 μ 初期値は 1、計算時間は地理的地図を初期値として計算した 1920 年エリアカルトグラム作成は CPU1.8GHz の計算機で 9.5 秒、一時点前のエリアカルトグラム上の地域形状を入力した他時点のエリアカルトグラム作成は全て 2 秒以内に完了している。

図-8 より、日本の総人口が増加すると同時に三大都市圏に人口が集中していく様子を視覚的に表現することができた。また、これらの図を用いてアニメーション表示することにより、人口の変化を効果的に視覚化することが可能である。アメリカ合衆国州人口および日本都道府県人口をエリアカルトグラムのアニメーションを用いて視覚化した例を下記の Web アドレスにて公開する。

<http://planner.tu-tokyo.ac.jp/gallery/carto/area.html>

5. おわりに

本論文では新たなエリアカルトグラムの作成手法を提案した。エリアカルトグラム作成問題を簡略化するため、地域を三角網に分割することを前提とし、各三角形の面

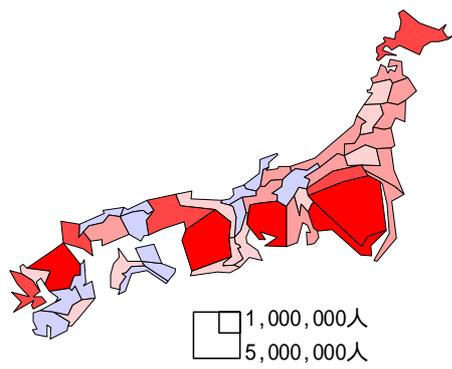
積を統計データに合わせるように三角網を変形する問題を設定した。その上で、統計データの表現に不必要な変形を抑え分かりやすいエリアカルトグラム作成を行うために、三角網を形成する各辺の方位角変化を抑える条件を設定し、不良設定問題であるエリアカルトグラム作成問題を正則化した。この解法の線形化を通して、線形最小二乗問題の繰り返しによりエリアカルトグラムを作成するアルゴリズムを構築した。また、三角網分割の方法として Delaunay 三角網を採用し、実データを用いて本提案手法の適用可能性を検討した。

本提案手法の特徴は、目的関数が簡潔かつ数学的に明快地記述されており、設定項目が少なく操作性に優れており、また短時間で計算できるという点である。これは、従来手法の問題点を解決した実用的な手法である。本提案手法は、GIS の拡張機能として容易に実装できる手法であり、将来的に一般の GIS ユーザーが統計データを簡単に視覚化できる環境を提供することができる。

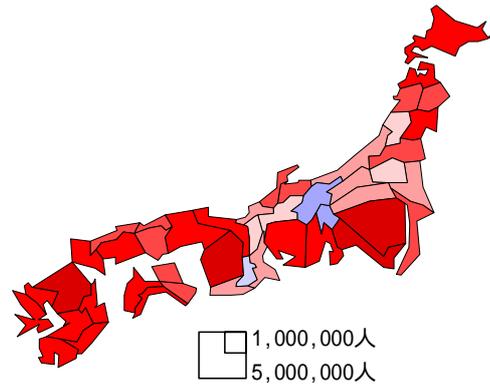
但し、本研究では従来手法との比較は行っていないのが現状である。連続エリアカルトグラム研究では共通の基準で表現精度・形状・計算時間を比較することが難しい。その要因は、例えば使用した地域形状のデータが異なる点、論文により表現精度の指標が異なる点等が挙げられる。また、エリアカルトグラム上の統計データの視認性の評価基準や地理的地図上とエリアカルトグラム上の地域形状の類似度の評価基準が論文によって異なる点も作成手法の比較が困難な理由の一つである。また、既往研究では作成アルゴリズムの過程が明快地記載されていないため、論文だけからその作成手法の再現を行うのが難しく、比較実験が行えないという問題もある。既往の作成手法を共通の基準で比較することは、エリアカルトグラム研究において今後の重要な課題である。

本提案手法では、簡略化された形状データを入力データとして用いている。形状データの自動簡略化手法の実装によって様々な対象のエリアカルトグラム作成が容易に行えるようになるほか、エリアカルトグラムへの詳細な海岸線を空間内挿することによってより印象的なエリアカルトグラム作成を行えると考えられる。

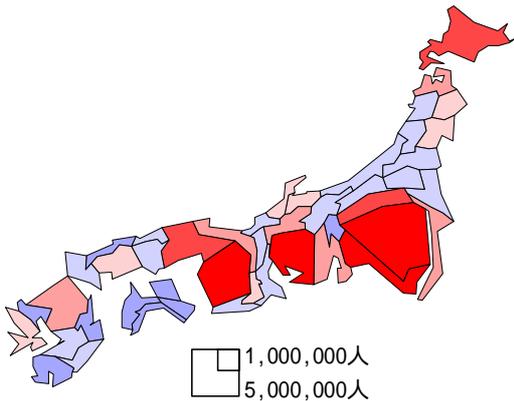
また近年、エリアカルトグラムは統計データの視覚化以外に、等人口密度空間などの特徴空間上における事象の空間分布を分析する手段¹⁴⁾としても用いられている。点や線のデータをエリアカルトグラム上に空間内挿することにより、例えば等人口密度空間上における道路網の配置の視覚化等にエリアカルトグラムを活用することも可能となる。このようにエリアカルトグラムを、統計データの視覚化および空間分析の手段としてより有効に活用するためには、空間内挿手法の整理・ツールの開発が不可欠である。これらを本提案手法の GIS 環境への実装と合わせて今後の課題とする。



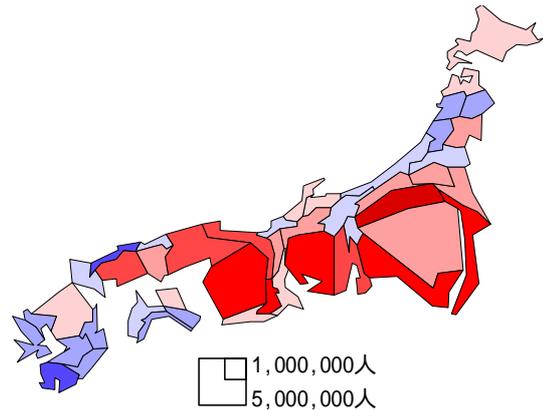
(a) 1940年



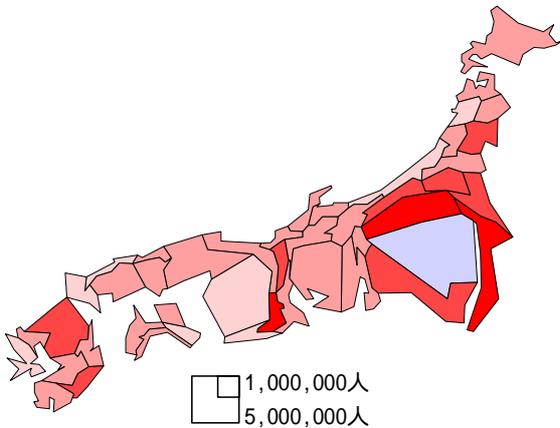
(b) 1950年



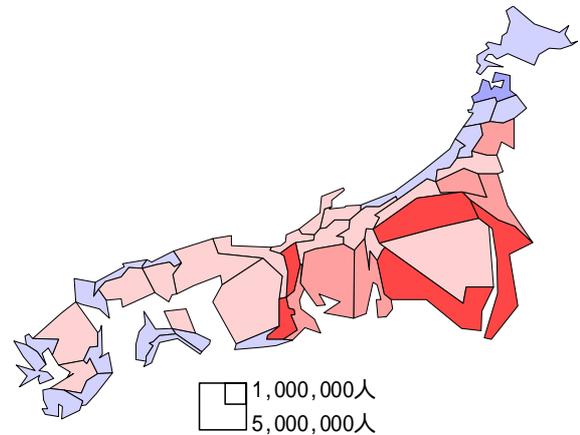
(c) 1960年



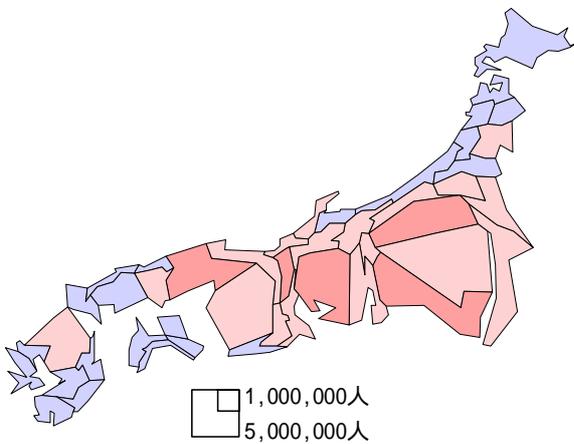
(d) 1970年



(e) 1980年



(f) 1990年



(g) 2000年

人口増加率(%)

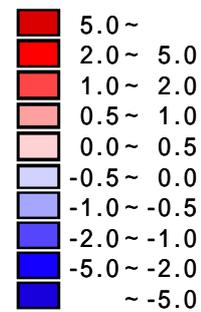


図-8 提案手法による都道府県人口・人口増加率エリアカルトグラム

謝辞：東京大学の森地茂教授(現名誉教授)には数多くの貴重な意見を頂いた。また，大学院生の Mongkol Tawechaitosapol 氏には計算処理等に協力を頂いた。ここに記して謝意を表する。

参考文献

- 1) 例えば大場亨：ArcView8で地域分析入門,成文堂,2003.
- 2) 例えばSlocum, T.A. : *Thematic Cartography and Visualization*, Prentice Hall, Upper Saddle River, 1998.
- 3) Tobler, W.R. : Thirty five years of computer cartograms, *Annals of the Association of American Geographers*, Vol.94, No.1, pp.58-73, 2004.
- 4) Tobler, W.R. : A continuous transformation useful for districting, *Annals of New York Academy of Science*, Vol.219, pp.215-220, 1973.
- 5) Tobler, W.R. : Pseudo-cartograms, *The American Cartographer*, Vol.13, pp.43-50, 1986.
- 6) Dougenik, J.A., Chrisman, N.R. and Niemeyer, D.R. : An algorithm to construct continuous area cartograms, *Professional Geographer*, Vol.37, No.1, pp.75-81, 1985.
- 7) Gusein-Zade, S.M. and Tikunov, V.S : A new technique for constructing continuous cartograms, *Cartography and Geographic Information Systems*, Vol.20, pp.167-173, 1993.
- 8) Kocmoud, C.J. and House, D.H. : A constraint-based approach to constructing continuous cartograms, *The Eighth International Symposiums on Spatial Data Handling Proceedings*, pp.197-204, 1998
- 9) Keim, D.A., North, S.C. and Panse, C. : CartoDraw: a fast algorithm for generating contiguous cartograms, *IEEE Transactions on Visualization and Computer Graphics*, Vol.10, No.1, pp.95-110, 2004.
- 10) Du, C. and Liu, K. : Constructing contiguous area cartogram using ArcView Avenue, *Proceedings of Geoinformatics '99 Conference*, pp.1-7, 1999.
- 11) 藤田宏, 今野浩, 田邊國士：最適化法, 岩波講座 応用数学, 岩波書店, 1994.
- 12) 岸本一男：領域の最適三角形群への分割アルゴリズム, 情報処理, Vol.19, No.3, pp.211-218, 1978.
- 13) 谷口健男：FEMのための要素自動分割 デローニー三角分割法の利用, 森北出版, 1992.
- 14) 例えば Selvin, S., Schulman, J. and Merrill, D.W. : Distance and risk measures for the analysis of spatial data: a study of childhood cancers, *Social Science and Medicine*, Vol.34, No.7, pp.769-777, 1992.

(2004.4.30 受付)

A NEW METHOD FOR CONSTRUCTING CONTINUOUS AREA CARTOGRAMS - VISUALIZATION TOOL FOR STATISTICAL DATA IN GIS ERA -

Ryo INOUE and Eihan SHIMIZU

A continuous area cartogram is a transformed map in which the map regions are resized relative to the geographic distribution of a data set. It is expected to be an effective visualization tool for statistical data in the GIS era.

The construction of continuous area cartograms is an ill-posed problem, for it is impossible to decide the region's shape by its size. Although many solutions have been proposed, none of these are mathematically clear, user-friendly, or computable in short calculation times.

We propose a simple, practicable, and fast algorithm which solves these exiting problems in this paper.