

三角網分割を用いた連続エリアカルトグラムの作成手法

井上 亮, 清水 英範

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その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。

キーワード：連続エリアカルトグラム、視覚化、三角網分割

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等の目的で利用されている。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、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。

他の視覚化手法に、計量地理学等で注目されている連続エリアカルトグラムと呼ばれる手法がある。連続エリアカルトグラムとは、地域の統計データの大小を地図上の面積の大小を用いて表現するように地理的地図を変形して統計データを表現する手法である。統計データの地理的分布を印象的に視覚化でき、読図者に地域の特徴を直感的に把握することを促す手法である。なお本稿では以後、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には今まで数多くの作成手法が

いのうえ りょう

東京大学 空間情報科学研究センター
〒277-8568 千葉県柏市柏の葉5-1-5

しみず えいはん

東京大学 大学院工学系研究科

〒113-8656 東京都文京区本郷7-3-1

提案されてきた。しかし、複雑なパラメータ設定・長時間の計算を要する、高精度でデータを表現できない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ソフトウェアが提供されている例はほとんど見られない。

エリアカルトグラム作成問題とは、地域の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3つの課題がある。

1. 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。

2. 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用いる際には、読図者は地理的地図との対比を通して地域形状の違いを認識し、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。

3. エリアカルトグラム作成手法をソフトウェアとして提供する場合には、パラメータ設定等に多くの試行錯誤をユーザー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、ユーザー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。

本稿では、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。

2. 既存のエリアカルトグラム作成手法

Tobler[1]が初めて計算機を用いたエリアカルトグラム作成法を提案して以来、様々な手法が提案されてきた。エリアカルトグラム研究の歴史についてはTobler[2]に詳しいが、ここではその概説を行う。

初期のエリアカルトグラム作成法は、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ため、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できなかった。しかし、Kocmoud・House[3]やKeim *et al.*[4]では、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カルトグラム作成が可能になっている。しかし、アルゴリズムは複雑で多くの初期値の設定が必要な上、出力カルトグラムの形状が初期値設定に強く依存するため、不適切な設定を行った場合、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り、使いやすい視覚化手法とは言えない。

一方、Gastner・Newman[5]は、数学的に明快な作成法を構築するという観点から、拡散方程式を応用した解法を提案している。複雑なパラメータ設定等を要さず高速に計算することができるが、初期の作成法の場合と同様に地域形状変形が大きくなり、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することはできない。

このように、既存の作成手法は視覚的に分かりやすいエリアカルトグラムの作成が可能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。

そこで、次節ではパラメータ設定等に複雑な手順を必要としない数学的に明快なエリアカルトグラム作成手法の構築を行う。

3. エリアカルトグラム作成手法の提案

3.1 基本的な考え方

前述のように、エリアカルトグラム作成問題は地域の面積の情報だけでは一意な地域形状が得られない不良設定問題である。しかし、分かりやすいエリアカルトグラム作成には地域形状変形を抑制する必要がある。そこで、地域形状変形を抑える正則化条件を設定し、これらの問題の解決を目指す。

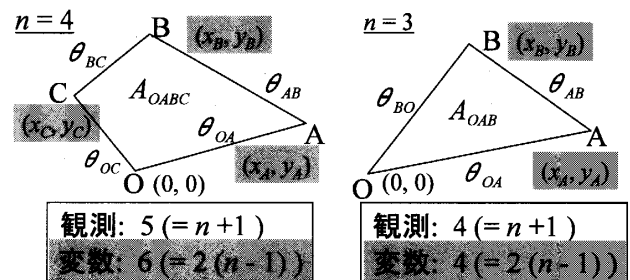
ここで、地域形状変形を抑制するため、地域の辺の方位角変化を抑える条件を導入することとする。もし方位角変化の抑制に対して極端に大きな重み付けを行うと、地理的地図からの拡大・縮小のみを許すことになり、形状を完全に維持した変形を行うことができる。しかし、辺が4本以上の多角形では、面積と各辺の方位角情報を与えても、形状を一意に定められない(図1)。しかし、三角形では観測方程式と未知変数の数が等しくなるため、形状を一意に定めることができる。

そこで地域を三角網分割し、三角網上の各辺の方位角を拘束する場合を考える。未知変数の数は、1点を座標系に固定するため、(頂点数-1)×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、地域形状を一意に求めることが可能になる。

そこで、本稿では地域を三角網分割することを前提とし、三角網上の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。なお、地域形状の分割はDelaunay三角網分割を用い、地域に与えられたデータを各三角形に面積按分する。

3.2 三角網分割を用いた作成手法の提案

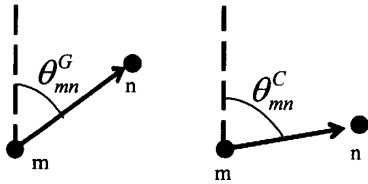
頂点 i, j, k からなる三角形 t_{ijk} のエリアカルトグ



ただし、 A_{OAB} : 多角形OABの面積

(x_A, y_A) : 頂点Aの座標、 θ_{OA} : 辺OAの方位角

図1 面積・方位角と頂点座標の関係



(a) 地理的配置上 (b) エリア
カルトグラム上

図2 辺 mn の座北方位角

ラム上の面積 A_{ijk} をデータ D_{ijk} に合わせる目的関数は、三角網中の三角形の集合を T とすると、式(1)と表せる。

$$\min \sum_{t_{ijk} \in T} (1 - A_{ijk}/D_{ijk})^2 \quad (1)$$

また、三角網上の辺の方位角変化を抑制する正則化項は、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} (図2)、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、式(2)と表せる。

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

主問題 (式(1)) と正則化項 (式(2)) より、エリアカルトグラム作成問題は、正則化項の重みを μ とすると、式(3)となる。

$$\min \{ \sum_{t_{ijk} \in T} (1 - A_{ijk}/D_{ijk})^2 + \mu \sum_{e_{mn} \in E} (\theta_{mn} - \theta_{mn}^G)^2 \} \quad (3)$$

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(3)を三角網の頂点座標を用いた式に改める。以降表記上の都合により、式(4)を用いる。

$$x_{ij} = x_j - x_i, \quad y_{ij} = y_j - y_i \quad (4)$$

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(5)となる。

$$A_{ijk} = |x_{ij}y_{ik} - x_{ik}y_{ij}|/2 \quad (5)$$

主問題 (式(3)第一項) に式(5)を代入する際に絶対値記号を取り除くため、主問題を式(6)と改め、

$$\min \sum_{t_{ijk} \in T} (1 - A_{ijk}^2/D_{ijk}^2)^2 \quad (6)$$

式(6)に式(5)を代入する。

$$\min \sum_{t_{ijk} \in T} \{ 1 - (x_{ij}y_{ik} - x_{ik}y_{ij})^2 / 4D_{ijk}^2 \}^2 \quad (7)$$

また、正則化項 (式(3)第二項) は、

$$+ \mu \sum_{e_{mn} \in E} \{ \arctan(x_{mn}/y_{mn}) - \arctan(x_{mn}^G/y_{mn}^G) \}^2 \quad (8)$$

と書けることから、

$$+ \mu \sum_{e_{mn} \in E} (x_{mn}/y_{mn} - x_{mn}^G/y_{mn}^G)^2 \quad (9)$$

とほぼ同値となる。さらに、計算上分母が0となるのを避けるため、式(10)とする。ただし、頂点 mn 間距離を d_{mn} と記述する。

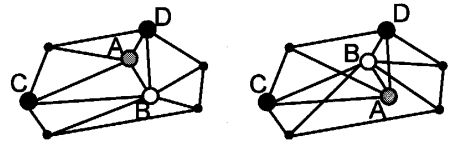


図3 $\triangle ABC \cdot \triangle ABD$ の反転と三角網の位相破壊

$$+ \mu \sum_{e_{mn} \in E} \left(\frac{x_{mn}y_{mn}^G - x_{mn}^G y_{mn}}{d_{mn}d_{mn}^G} \right)^2 \quad (10)$$

以上より、エリアカルトグラム作成問題は式(11)と書ける。

$$\min \left[\sum_{t_{ijk} \in T} \left\{ 1 - \frac{(x_{ij}y_{ik} - x_{ik}y_{ij})^2}{4D_{ijk}^2} \right\}^2 + \mu \sum_{e_{mn} \in E} \left\{ \frac{x_{mn}y_{mn}^G - x_{mn}^G y_{mn}}{d_{mn}d_{mn}^G} \right\}^2 \right] \quad (11)$$

式(11)の非線形最小二乗問題を解けば、エリアカルトグラムが作成できる。ただし、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、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(11)を近似値近傍で線形化すると、

$$\min \left[\sum_{t_{ijk} \in T} \left\{ 1 + \frac{3A_{ijk}^2}{D_{ijk}^2} + \frac{A'_{ijk}}{D_{ijk}^2} (y'_{jk}x_i - y'_{ik}x_j + y'_{ij}x_k - x'_{jk}y_i + x'_{ik}y_j - x'_{ij}y_k) \right\}^2 + \mu \sum_{e_{mn} \in E} \left\{ \frac{(x_n - x_m)y'_{mn} - x'_{mn}(y_n - y_m)}{d_{mn}^2} \right\}^2 \right] \quad (12)$$

ただし、 $A'_{ijk} = (x'_{ij}y'_{ik} - x'_{ik}y'_{ij})/2$

と表せる。式(12)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。しかし、式(12)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止
- 実用的な計算時間で作成するため、解法を高速化

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。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3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(12)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）が大きくなる。そのため式(12)は、最適解近傍では収束速度が低下してしまう。この方法を高速に収束させるためには、最適解近傍で $\mu \rightarrow 0$ となるよう制御する必要がある。

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

$$\left\{ \frac{\sum_{t_{ijk} \in T} (D_{ijk} - A_{ijk})^2}{\sum_{t_{ijk} \in T} (D_{ijk} - A'_{ijk})^2} \right\} \mu \quad (13)$$

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。

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。

4.1 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、頂点数を185に簡略化したアメリカ合衆国地図（図4）を用いる。ただし、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。

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。

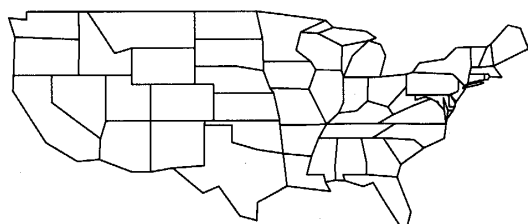
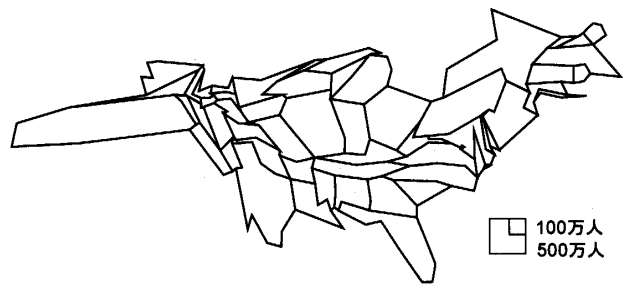


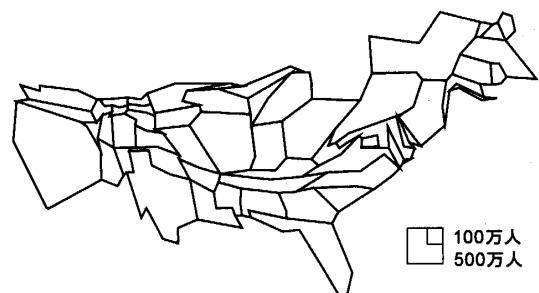
図4 入力したアメリカ合衆国州形状データ

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、図5は人口データをほぼ完全に表現していることが分かる。

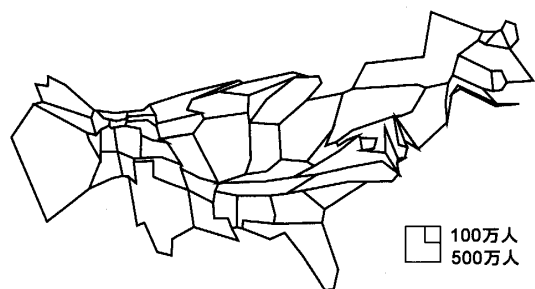
次に、 μ 初期値による出力エリアカルトグラムの違いについて注目する。 μ 初期値が小さいと、収束速度が速く計算時間が短い（表1）、形状変形は大きくなり分かりにくいカルトグラムを出力する（図5）。



(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

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	1	$\sqrt{10}$	10
収束計算(回)	19	28	101
計算時間(秒)	5.0	7.5	23.8
RMSE(人)	25.5	2.1	8.3

(ただし、CPU2.8GHzの計算機使用)

表2 1980年アメリカ人口エリアカルトグラム
ヘルマート変換自由度修正済み決定係数

μ 初期値	1	$\sqrt{10}$	10	
国形状	0.907	0.917	0.908	
州形状	最小値	0.170	0.895	0.886
	中央値	0.998	0.999	0.999
	標準偏差	0.141	0.022	0.020

ここで、カルトグラム上形状と地理的形状間の類似度を定量的に評価することを試みる。各形状の頂点を合わせるようにヘルマート変換を行い、その自由度修正済み決定係数を2形状の類似度指標として用いる(表2)。まず国全体の形状を評価した場合、大きな値の違いは見られない。この要因の一つに、アメリカ合衆国は東海岸に小さな州が集中し頂点が多いため、東海岸の形状が大きく評価され、西海岸の形状の変形を反映できていないことが挙げられる。そのため、形状を評価するには頂点の分布などに注意を払う必要がある。そこで州毎に形状比較を行い、 μ 初期値1の場合、西海岸3州(カリフォルニア・オレゴン・ワシントン)で著しい変形を起こすことを確認した。一方、 μ 初期値が $\sqrt{10} \cdot 10$ では、値に顕著な違いは見られず、ほぼ同等の形状を維持していることが確認できる。

以上のように、 μ に大きな初期値を設定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、形状変化を抑えかつ高速に計算できる 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。

4.2 視覚化例の提示—都道府県人口データ

総務省統計局整備の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。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。

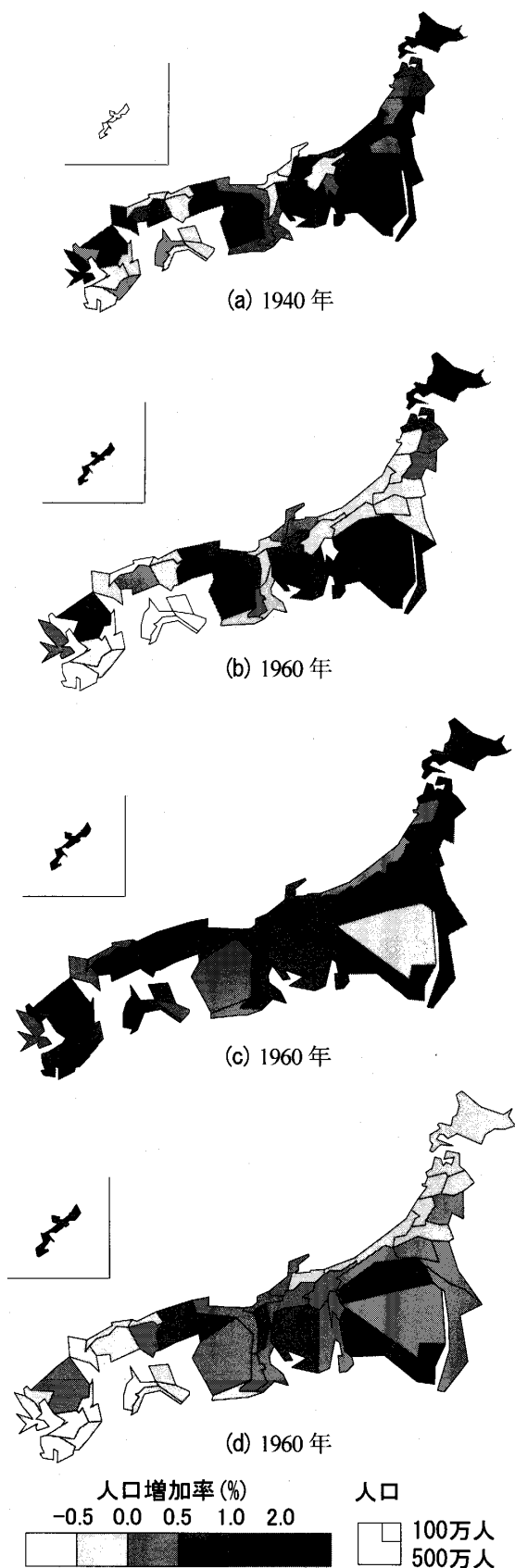


図6 提案手法による都道府県人口・人口増加率
エリアカルトグラム

その中から1940~2000年の20年毎の都道府県人口エリアカルトグラムを示す(図6)。ただし、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CPU 2.8 GHzの計算機で8秒、一時点前のエリアカルトグラム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秒以内に完了した。

図6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、これらの図をアニメーション表示することにより、人口変化を効果的に視覚化できる。人口の変遷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アドレスにて公開する。

<http://home.csis.u-tokyo.ac.jp/~inoue/research/cartogram/>

5. おわりに

本稿ではエリアカルトグラムの作成手法を提案した。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、地域を三角網に分割し、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すると同時に、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て、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。その上で、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、実データを用いて本提案手法の適用可能性を検討した。

提案手法の特徴は、目的関数が簡潔かつ数学的に明快に記述されており、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。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。

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。

なお、本稿は、筆者らの先行論文[6]をもとに一部を再構成し、本特集号向けに要約したものであることを付記しておく。

参考文献

- [1] Tobler, W. R. (1973) A continuous transformation useful for districting, *Annals of New York Academy of Science*, 219, 215-220.
- [2] Tobler, W. R. (2004) Thirty five years of computer cartograms, *Annals of the Association of American Geographers*, 94 (1): 58-73.
- [3] Kocmoud, C. J. and House, D. H. (1998) A constraint-based approach to constructing continuous cartograms, *The Eighth International Symposium on Spatial Data Handling Proceedings*, 197-204.
- [4] Keim, D. A., North, S. C. and Panse, C. (2004) CartoDraw: a fast algorithm for generating contiguous cartograms, *IEEE Transactions on Visualization and Computer Graphics*, 10 (1), 95-110.
- [5] Gastner, M. T. and Newman, M. E. J. (2004) Diffusion-based method for producing density-equalizing maps, *Proceedings of the National Academy of Sciences of the United States of America* 101 (20), 7499-7504.
- [6] 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, (779/IV-66), 147-156.