

非階層型クラスタリングを用いた地域の状態遷移の類型化手法の提案 ～関東地方一都六県を対象として～

寺木彰浩

Area classification of status transition by non-hierarchical clustering

– In case of KANTO region –

Akihiro TERAHI

Abstract: This paper deals with a method to classify status transition of area. A series of grid cell data is available in Japan. It is very useful for local governments and people to find a similar state area in other area. Our method uses non-hierarchical clustering. A case study is done for KANTO region and shows urbanization process in that area.

Keywords: 標準地域メッシュ (third standard grid cell), 状態遷移 (status transition), 可視化 (data visualization)

1. はじめに

人口減少や大型店の出店拡大、車社会の発展により人々は郊外の大規模駐車場を完備した大規模ショッピング施設に流れ、地方都市の中心市街地は衰退しつつあるといわざるを得ない。それぞれの自治体や住民は自ら置かれた状況に応じた対応を求められている。しかし俯瞰的に似通った状態の地域を参考にするすることで、いわゆる「車輪の再発明」に力を削がれることなく本当に必要な事項への注力や、事例を通じた知見・経験の蓄積により更に進んだ対応が可能になるはずである。

類似した状況の地区を探すにあたり、現状だけでは十分ではない。時間経過に伴う一連の状況の変遷を類型化することにより、現在の状況に似通った過去の他の地区を探し出し、その変遷を追うことで元の地区の将来を推測する可能性がある。

本研究は地域メッシュデータを対象にクラスタ分析を実施し、類型化したメッシュの状態遷移を地理情報システムにより可視化する手法について提案するものである。

なお本稿は根本嗣美氏の千葉工業大学卒業論文

寺木彰浩：〒 275-8588 習志野市津田沼 2-17-1

千葉工業大学 創造工学部 都市環境工学科

Email: teraki.akihiro @ it-chiba.ac.jp

を筆者の責任でとりまとめた。

2. 地域メッシュによる分析の概要

(1) 地域メッシュについて

日本では国土利用計画の策定などを目的として 1960 年代から様々なデータが蓄積されてきた。特に地域メッシュ統計は、昭和 48 年 7 月 12 日行政管理庁告示第 143 号に従って、緯度・経度に基づき地域を隙間なく網の目の区域に分けて、それぞれの区域に関する統計データを編成したもので以下のような利点を持つ（総務省統計局，2019）。

- ①ほぼ同一の大きさ及び形状の区画が単位であり、相互の計量的比較が容易
- ②位置や区画が固定されており、地域事象の時系列的比較が容易
- ③位置の表示が明確で簡便なので、距離などに関連した分析、計算、比較が容易

今回は関東一都六県を対象に、それらをカバーする範囲ではメッシュの大きさの違いは無視できるほど相対的には小さいとみなして、3 時点の時系列データを用いてクラスタ分析を行った。

(2) 使用データ

(I) 関東一都六県のデータ収集

地域住民が容易に入手できるデータを基に、類

表-1 データの種類と年次

データの種別	95 年	05 年	15 年
国勢調査	1995 年	2005 年	2015 年
商業統計メッシュデータ	1994 年	2004 年	2014 年
土地利用	1997 年	2006 年	2014 年

表-2 データを収集した範囲

メッシュコード	38	39	40
52	5238	5239	5240
53	5338	5339	5340
54	5438	5439	5440
55	5538	5539	5540

表-3 土地利用の区分

分類	細目
農地	田, その他農用地, ゴルフ場, その他用地
未利用地	森林, 荒地
都市	建物用地, 道路, 鉄道
水面	河川湖沼, 海浜, 海水域

似した状況の地区を自ら探し出すことを念頭に、以下のデータを収集した。各メッシュの状態遷移を把握するため、国勢調査の年次を基に表-1 に示すデータを収集した。地理的範囲を表-2 に示す。

・人口

本研究では都市化の程度を示す指標として用いている。e-Stat 政府統計の総合窓口で公開されている国勢調査の人口総数を用いた。

・産業

栃木県、群馬県、茨城県、埼玉県は平成 13 年～27 年までの市町村内総生産と産業別生産額のデータを公開している。東京都、神奈川県、千葉県は公開していない。事業所企業統計と経済センサスのデータは e-stat で公開されている。また商業統計メッシュデータは経済産業省のホームページに平成 26 年度のみ公開されている。他年度の調査データは経済産業調査会統計情報部にて販売している。このうち本研究では商業統計メッシュデータの小売業計の年間販売額（以下、販売額）を用いた。

・土地利用

国土数値情報ダウンロードサービスにより土地利用 3 次メッシュデータが公開されている。本研究では表-3 のように分類し、その合計値を使用した。

(3) 用いた手法

(I) クラスタリング

各メッシュの属性を類型化する際に R でクラスタリングを行った。クラスタリングは

・階層的手法：類似度などに基づき逐次併合

・非階層的（分割最適化）手法：評価関数を設定し、その値を最も良くする分割を探索

あるいは

・ソフトクラスタリング：複数のクラスタに属することを許容する。

・ハードクラスタリング：複数のクラスタに所属することを許容しない。

に大別される。本研究では非階層的かつハードなクラスタリングである k-means 法と x-means 法を用いた。

k-means 法はクラスタの平均を指標として、あらかじめ与えられたクラスタ数 k 個に分割するものである。x-means 法はクラスタ数を変えながら k-means 法を繰り返し適用し、BIC (Bayesian Information Criterion) を基準として妥当と判断されるまで分割を繰り返すものである（石岡, 2000）。

k-means 法のクラスタ数を与える基準はいくつか知られているが本研究では R に実装されている GAP 統計量を用いた（Tibshirani et al., 2002）。

クラスタリングを行う際に年次も属性としたので同一メッシュが 3 時点で重なるため、3 重にサンプルとなっている。

この工夫により、同一の区分で 3 時点のメッシュが個別に分類されるので状態遷移の把握が容易となるが、データ量が増加し、計算負荷や記憶負荷が無視し得ないものとなる。

(II) 可視化

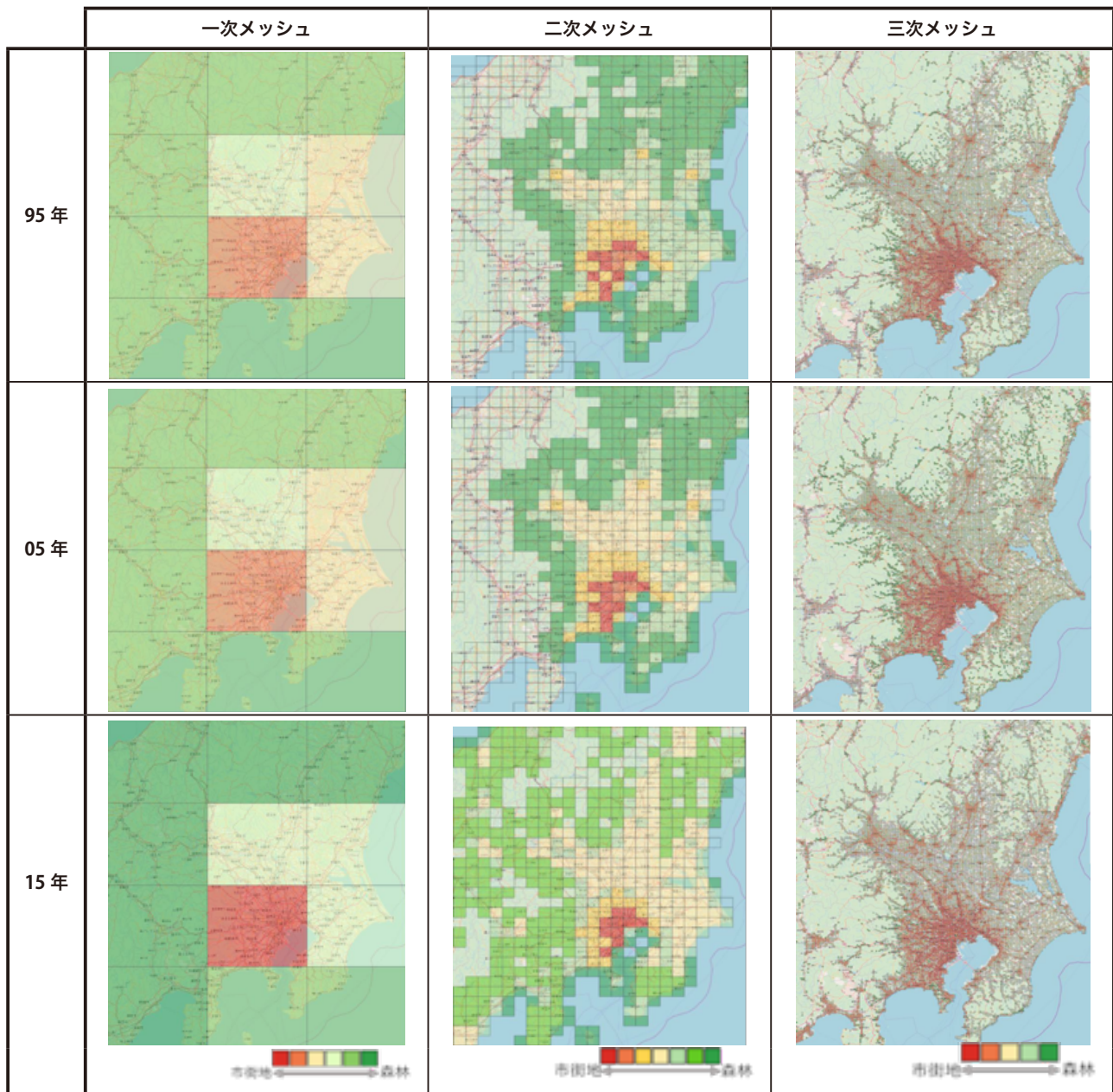
前項の結果を QGIS 上で図化した（表-4）。メッシュコードによる位置参照を処理するためプラグインの JapanMesh を使用した。

3. 結果の概要

(1) 前処理

多重共線性を考慮し、土地利用の水面を抜いた計 9 つの指標で相関係数を計算した。その結果、

表-4 クラスタリング結果の図示



農地、未利用地、人口、販売額の4指標を用いてクラスタリングを行うことにした。

また、それぞれのデータはマハラノビスの汎距離により標準化した。

(2) メッシュレベルごとの結果 (表-4)

(I) 一次メッシュ (表-5)

x-means 法と GAP 統計量共にクラスタ数6となったためクラスタ数を6としてk-means 法を実施した結果を用いた。各クラスタの中心の値から市街化⇔森林の程度を判断し、QGISにより図示した。

(II) 二次メッシュ (表-6)

x-means 法の結果、クラスタ数は65となったが、クラスタ数が多すぎて意味がない。また GAP 統計量ではk=1であった。

そのため3次と同じクラスタ数5を基準とし、クラスタ数5～9全てでk-means 法を実行した。その結果を各クラスタの散布図に表示し、使用する分析結果を決定した。

クラスタ数6の時は、クラスタ数5の時よりもすべての指標でより詳しく区分された。次にクラスタ

表-5 一次メッシュの結果

番号	農地	未利用	人口	販売額
1	0.24	2.09	-0.33	-0.37
2	0.56	-0.46	0.13	0.07
3	0.60	-0.40	3.11	4.00
4	0.47	0.47	3.50	-0.33
5	-0.82	-0.55	-0.42	-0.31
6	1.89	-0.07	0.07	-0.02

表-6 二次メッシュの結果

番号	農地	未利用	人口	販売額
1	-0.21	-0.65	4.46	2.30
2	0.49	-0.31	-0.08	-0.06
3	0.56	-0.54	1.60	1.11
4	-0.36	1.86	-0.36	-0.28
5	-0.77	-0.80	6.39	11.22
6	1.77	-0.32	-0.08	-0.12
7	-0.85	-0.57	-0.34	-0.22

表-7 三次メッシュの結果

番号	農地	未利用	人口	販売額
1	0.11	0.72	-0.49	0.14
2	-1.19	-0.66	2.47	0.82
3	-0.71	-0.50	0.26	0.02
4	-0.98	2.33	0.55	-0.16
5	1.05	-0.48	0.45	-0.12

数7の時は、人口総数がより詳しく区分され、クラスタ数8と9の時は、農地がより詳しく分類された。本研究では「市街化が進んでいるか」という点に焦点を当てた地域分類を行うため、農地を詳しく見る必要はないと判断した。そのため、二次メッシュはクラスタ数7を採用した。

(Ⅲ) 三次メッシュ (表-7)

3時点を通じてデータがあったメッシュを対象とした。全部で88,083（毎期29,361）である。

GAP統計量によりクラスタ数を5としてk-means法を実行した。x-means法では133である。

図-1は都心部の、図-2は北関東の95年と15年の比較を表している。クラスタ4→5→1→3→2の順に都市化が進んでいると判断し、8段階で評価した。赤に近づくほど都市化が進み、緑はその逆である。また、透過率が高いメッシュは変化がなかった地域であり、中抜きされているメッシュは3時点のデータが揃っていない地域である。水面と、どの年度もデータが欠損、あるいは非公開となっている地域はメッシュが描かれていない。

都心部は変化している地域は少なかったが、埼玉県の越谷市と戸田市は都市化が進んでいた。ま

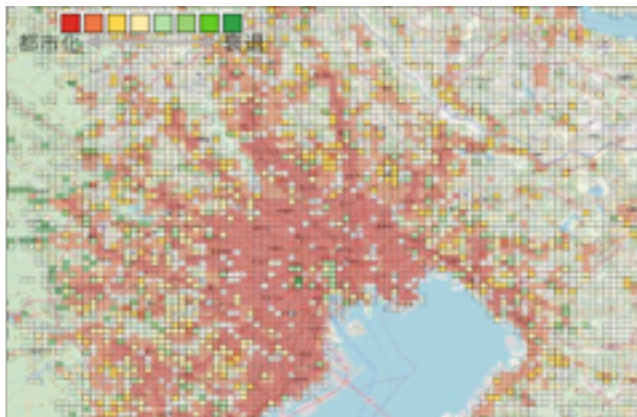


図-1 都心部の変遷

た、他の変化があった地域においても、比較的都市化の傾向にある地域が多かった。しかし、都内の中央区と千代田区周辺は衰退しており、関東の中で最も衰退傾向にある地域は原宿という結果が得られた。明治神宮と代々木公園の影響と思われる。

宇都宮駅の周辺はクラスタ3に分類されるが、3時点での変化を見ると衰退の傾向にあった。群馬県の前橋市と高崎市も同様の傾向が見られ、北関東は衰退していると判断できる。

4. 今後の課題

2.(3)(i)で述べた計算負荷や記憶負荷は関連技術の進展で軽減されるであろう。クラスタ間の遷移をどのように集約してわかりやすく表現するかが重要な課題になると考える。

参考文献

- 1) 根本嗣美, 2019, 非階層型クラスタリングを用いた地域類型化手法の開発ー関東地方一都六県を対象としてー, 千葉工業大学卒業論文.
- 2) 総務省統計局, 地域メッシュ統計の特質・沿革, <https://www.stat.go.jp/data/mesh/pdf/gaiyo1.pdf>, 2019年8月20日参照.
- 3) 石岡恒憲, 2000, クラスタ数を自動決定するk-meansアルゴリズムの拡張について, 応用統計学, **29**(3), 141-149.
- 4) Tibshirani, R., Walther, G. and Hastie, T. 2002, Estimating the number of clusters in a data set via the gap statistic, *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **63**(2), 411-423.

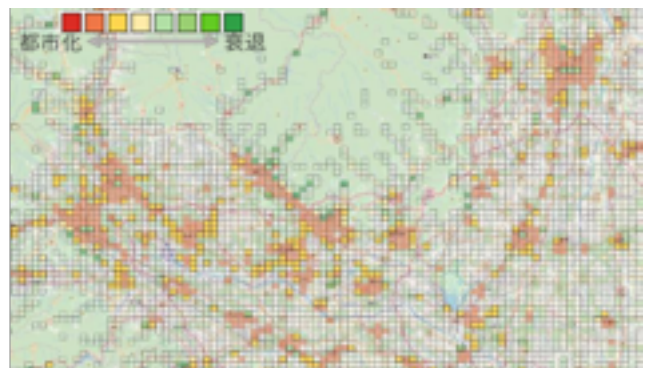


図-2 北関東の変遷