



ESRI Research Note No.78

政令市千葉市を事例にした小地域の特徴を 機械学習で理解する一手法

村舘 靖之

May 2023



内閣府経済社会総合研究所

Economic and Social Research Institute
Cabinet Office
Tokyo, Japan

ESRI Research Note は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません（問い合わせ先：<https://form.cao.go.jp/esri/opinion-0002.html>）。

ESRI リサーチ・ノート・シリーズは、内閣府経済社会総合研究所内の議論の一端を公開するために取りまとめられた資料であり、学界、研究機関等の関係する方々から幅広くコメントを頂き、今後の研究に役立てることを意図して発表しております。

資料は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません。

The views expressed in “ESRI Research Note” are those of the authors and not those of the Economic and Social Research Institute, the Cabinet Office, or the Government of Japan.

政令市千葉市を事例にした小地域の特徴を機械学習で理解する一手法

村館靖之¹

要旨

ゾーニングを行う土地計画・都市計画等、自治体において小地域ごとに一定の類似性を見出し、開発の方向性の立案や施策を行う業務は多く存在する。本稿では新しい類似性の抽出手法として小地域クラスタリングの手法を提案する。小地域クラスタリングとは、地理情報データを基に類似した性質の町丁をグループにまとめることを指す。国勢調査から2005年、2010年、2015年、2020年の4つの時点について、千葉市の小地域クラスタリングを行い、変遷を追った。自治体の政策担当者との対話を通じ、分析に追加すべき変数等を特定し、都市計画、防災計画等への応用などを目指したい。

もくじ

1. はじめに	3
2. データおよび手法	4
3. 関連する先行研究	7
4. 分析結果	10
5. 考察	14
6. 結論	15
参考文献	16
付録	17

1. はじめに

都市計画など、地区ごとに一定の類似性を見出し、開発の方向性の立案や施策を行う業務は多く存在する。特定の指標に基づく地区ごとの特性は既存の統計や地図情報から入手可能であるが、複数の指標による総合的な特性の抽出は自治体の政策担当者の経験と勘に

¹ 村館靖之：内閣府経済社会総合研究所研究専門職

本稿の作成にあたって、千葉市総合政策局総合政策部政策企画課統計室の職員の方から有益なコメント、助言をいただいた。また2022年3月の情報処理学会全国大会(愛媛大学)において、座長の後藤晶先生から有益なコメントをいただいた。2023年3月の社会情報学会関東支部研究会において、電気通信大学の山本佳世子先生、岩本茂子先生から有益なコメント、指導をいただきました。内閣府の野村裕官房審議官、内閣府経済社会総合研究所の増島稔所長、林伴子次長、中澤信吾総括政策研究官、桑原進総括政策研究官、金沢学院大学の貫裕二先生から、執筆にあたり詳細なコメントや指導をいただいた。ここに謝意を示したい。なお残りうる過ちは全て筆者の責任である。

基づく部分が多い。地区の開発方針を検討する上では、その地区に住む住民の総合的な特徴も一定の役割を果たせると考えられる。データに基づく政策立案の方向に向かって移行し、データ分析を政策に一層活用する方法の一つとして小地域クラスタリングの手法を提案する。クラスタリングは、様々な場面で用いられるが、例えばマーケティングでは顧客をデータに基づき分類する手法である。同一のグループに属する顧客に類似した商品を推奨する際になどに用いられる。本稿でいう小地域クラスタリングとは、国勢調査のデータから、類似した性質の町丁をグループにまとめることを指す。地図上で統計情報を可視化する際に、付加的情報として政策の単位（町丁レベル）の地域特性を示すことで、データに基づく政策立案の新しい手法を提案する。

はじめて地域を担当する新任の担当者や、十分経験を積んでいる担当者であっても、地域の特徴をデータに基づいて知る、再認識するための一手法としても、本稿で提案する小地域クラスタリングの手法は有用であろう。

国勢調査は人口の観点から、まち、ひと、しごとについて集計された基本的な統計で町丁ごとの小地域集計を含めて、オープンデータとして利用が可能であり、アクセスが容易である。地区の住民の特性という観点では納税情報（＝所得情報）や健康に関する情報（健診結果やレセプト）なども利用できると追加的分析の可能性はあるが、データ利用に向けて克服すべき課題は多く、自治体内部や内部であっても分析は難しいと言える。

千葉市の政策担当者と議論したところ、クラスタリングの際に時系列で結果を比較して、成長ないし衰退傾向にある町丁に関する情報を参考にしたいとの意見を得た。これを受けて、2005年、2010年、2015年、2020年の国勢調査についてクラスタリングした結果を比較する。

2. データおよび手法

2-1. データ

国勢調査の GIS データセットで境界情報と統計情報（2005年版、2010年版、2015年版、2020年版）を入手する²。対象を千葉市に設定している。千葉市は東京都からの通勤圏内の政令市である。国勢調査の GIS データセットでは、男女別人口総数及び総世帯数、年齢（5歳階級・4区分）、男女別人口・所有の関係別一般世帯数、住宅の建て方別世帯数、産業（大分類）および従業上の地位別就業者数、職業（大分類別）就業者数、世帯の経済構成別一般世帯数の項目が利用できる³。国勢調査は自治体等が申請すれば個票も利用できる

² データは地図で見る統計(統計GIS)サイト (<https://www.e-stat.go.jp/gis>) から入手した。最終訪問(2022/12/08)。

³ 毎年のデータ項目数は2005年121項目、2010年123項目、2015年126項目、2020年131項目である。ここから、年少人口比率、生産年齢人口比率、高齢化率、第一次産業従事者比率、第二次産業従事者比率、第三次産業従事者比率、ホワイトカラー比率、グレーカラー比率、ブルーカラー比率を取り出して使った。

るが、まずはオープンデータであり、最小の地域のまとまりの一つである町丁（小地域）単位の集計表を分析することを試みる。国勢調査のデータを利用することに関しては、行政オープンデータであることがメリットである。また人口、世帯、仕事といった自治体に関する基本情報が分析できる。他にも GIS ソフトを使って可視化することを想定すると、経済センサスや自治体内部の分析目的用途であれば KDB（国保データベース）の掃き出しデータを分析に用いることが考えられる。

分析にあたっては、人口密度の変数を追加し、特徴量を絞るため、緯度経度の他に、年少人口比率、生産年齢人口比率、高齢化率、第一次産業従事者比率、第二次産業従事者比率、第三次産業従事者比率、ホワイトカラー比率、グレーカラー比率、ブルーカラー比率を使った。変数の選択により、結果は異なると思われるが、本稿では、人口学的特徴、就業に関わる特徴を中心に変数を選択した。

また GMM（混合ガウスモデル）で分析するにあたっては、事前にデータを正規化して分析した。

2-2. 分析手法

分析手法として GMM（混合ガウスモデル）を採用する。この手法の特徴はデータからクラスタリングする際に、最適クラスタ数の情報について BIC（ベイズ情報量基準）を比較して活用できることだ。K-means 法も類似の手法であるが、この手法はクラスタ数を決め打ちしており、どのクラスタ数が最適であるかの情報は与えてくれない。

まずクラスタ分析の考え方を比較、整理したい。N 個のデータ集合（信号とも呼ぶ）が与えられた時に、近い属性のデータをまとめて、全体を m 個の部分集合に分割する。この m 個の部分集合をクラスタと呼ぶ。

クラスタというまとまりに分類するアルゴリズムには、階層的クラスタリングと非階層的クラスタリングの 2 種類が存在する。階層的クラスタリングは樹状図（デンドログラム）のように階層性を持ったデータ構造を分析し、クラスタに分類するアルゴリズムである。階層的クラスタリングは比較的小さいデータの分類に対して有効な手法である。

非階層的クラスタリングは、データの階層構造を仮定せず、クラスタに分類するアルゴリズムである。非階層的クラスタリングは比較的データ数が大きい場合の分析にも適しており、あらかじめ分類クラスタ数(k)を指定して分類する k-means(k 平均法)が代表的である。

k-means のアルゴリズム⁴

1. 初期段階：初めに m 個のクラスタ中心 η_1, \dots, η_m についてお互いを異なるように勝手

⁴ k-means のアルゴリズムの説明に関しては、甘利（2014）p.138 の記述を基に執筆した。k-means のアルゴリズムの問題点としては初期値依存で局所最適解におちいってしまうおそれが挙げられる。

に選ぶ。

2. 分割段階：データ集合Dの各要素 x_i を順番に取り上げ、これがどのクラスタ中心が一番近いかを計算し、一番近いクラスタ C_h に入れる。

$$x_i \in C_h: D_\phi[x_i: \eta_h] = \min_k D_\phi[x_i: \eta_k]$$

($D_\phi[x_i: \eta_k]$ はデータ x_i とクラスタ中心 η_k のダイバージェンス (凝距離))

3. 更新段階：各データがどのクラスタに属するか定まった後に、その構成データの平均として各クラスタ C_h の中心をもとめ、これを更新したクラスタ中心 η_1, \dots, η_m とする。
4. 終了段階：上記を繰り返し、収束した時点で終了する。

k-means ではあらかじめクラスタ数 k を仮定する必要があるが、最適なクラスタ数を探索するアルゴリズムとして、GMM (混合ガウスモデル) や x-means がある⁵。GMMではBIC(ベイズ情報量規準)に基づき、モデルを比較して、最適なクラスタ数を探索する。

BIC(ベイズ情報量規準)⁶:

$$\ln p(D) \approx \ln p(D | \theta_{MAP}) - (1/2)M \ln N$$

ベイズ情報量規準は上記の式で近似される。つまりあるデータ集合Dに対して確率分布 $p(D)$ が定まり、その尤度 $\ln p(D)$ は事後分布最大化推定値 (ベイズ推定値) θ_{MAP} で評価した尤度 $\ln p(D | \theta_{MAP})$ から M (パラメーター数)、 $\ln N$ (データ数の対数值) からなる $(1/2)M \ln N$ を引いたもので近似される。

k-means ではあらかじめクラスタ数 k を決め打ちしていたが、GMM という手法はBICによりモデルのフィット具合を評価し、最適なクラスタ数を探索することができる。このため、GMM の手法を用いることで、最適クラスタ数を評価した上で、分類することが可能となる。

行政担当者が、機械学習やクラスタ分類をする前の統計データの段階で、GIS (地理情報) ソフトを用いて、統計データ、行政データを地図上に可視化することも可能で、十分に意義がある。しかし、例えば GIS ソフトを使って地図上に可視化すると、人口規模のように一度に一つの項目について、主題図として可視化することになる。複数のデータを圧縮してまとめて地域属性を可視化する手法として、クラスタリングの手法と GIS ソフトの併用が考えられる。ただ一種類のデータを地図上に可視化するのではなく、複数の、

⁵ 国勢調査の小地域集計のデータのクラスタリングに関してはデータ数の観点から、大きいデータ数に適した非階層クラスタリング、クラスタ数を決め打ちではなく最適なクラスタ数を探索する観点から GMM (混合ガウスモデル) を分析手法に選択した。

⁶ BIC の近似式については Bishop. (2006). Ch. 4. pp. 216-217 を参照。BIC はモデルのフィット具合を評価するための式である。式: $\ln p(D) \approx \ln p(D | \theta_{MAP}) - (1/2)M \ln N$ の両辺に -2 を掛けて、 $BIC = -2 \ln p(D | \theta_{MAP}) + M \ln N$ としている文献 (赤池他(2007), p. 67) もあるが、本質的な違いではない。なお、モデル選択の罰金項 $\ln N$ を 2 とおくと AIC (赤池情報量規準) となる。

大規模なデータを圧縮して地図上に可視化する手法がクラスタリング手法と GIS ソフトの併用である。

本稿の分析手法の利点としては、行政の情報を単に GIS ソフトによって地図上に可視化するだけでなく、分析処理をすることで、複数の地域属性をまとめたクラスタ属性という付加価値を追加した可視化が可能になる。

小地域単位であれば、自治体の経験を積んだ行政担当者はまちの個性を把握している可能性が高い。しかし、同一性の高い小地域についてグループ化を行い、予備知識がなくとも新しく担当になった地域の個性を統計から抽出するために、本稿の手法は役立つのではないか。また同じ政令市の中の離れた行政区の小地域の対であっても、同じクラスタに分類されることがある。つまり政令市全体で最適な施策を立案するにあたって、本研究の手法は重要といえるのではないか。

クラスタ分類はアルゴリズムで自動的にできるが、結果の解釈は分析者に委ねられている。できるだけ標準化した手法をマニュアル化すべきといえる。

分類したクラスタを被説明変数にして、その他の変数を全て説明変数にして決定木分析にかけることで、クラスタごとの分岐に関わる説明変数を抽出する。決定木分析の結果を参考してクラスタごとの集計表を作り、分類されたクラスタの特徴を探る。

地理情報システムソフトとしてはフリーソフトの MANDARA を採用する。MANDARA の利点はウィンドウズ端末上でしか利用できないものの、フリーソフトとして利便性が高い点にある。

3. 関連する先行研究

関連する先行研究としては、桐村(2007)は k-means だけでなく、自己組織化マップの手法を用いて小地域のクラスタリングを行う手法を検討している。李(2009)は 2000 年の大分県別府市(人口 12 万人)の国勢調査データをもとに、自己組織化マップと階層的クラスタリングの Ward 法および非階層的クラスタリングの k-means 法で分類・分析を行っている。自己組織化マップはニューラルネットワークを応用し、次元削減する(=データを圧縮し、多数の特徴量から少数の特徴量にデータの次元を絞る)手法であり、本研究の手法である GMM(混合ガウスモデル)⁷と代替的な研究手段である。

佐藤(2011)は k-means および x-means で東京大都市圏をクラスタリングした先行研究である。x-means は GMM と同様、クラスタ数の最適値をベイズ情報量基準について比較することで求めるアルゴリズムである⁸。本稿は、地方自治体の政策立案に焦点を当て、国勢調

⁷ GMM(混合ガウスモデル)は機械学習の文脈で、複数の正規分布の重なりを仮定して、クラスタに分類するアルゴリズムを意味する。k-means 法、GMM の解説に関しては Bishop, (2006). Ch. 9, pp. 423-459 を参照。

⁸ x-means 法は Pelleg & Moore (2000)によって開発された k-means 法を拡張した手法で、k=2 から再帰的に k-means 法を適用し、クラスタ分割の結果を BIC(ベイズ情報量規準)

査のデータに関して GMM の手法を用いて分析するところに特徴がある。

浅川(2008)は本稿の問題意識と近い研究である。階層的クラスタリング(例:Ward法)やk-means法では最適クラスタ数が決定されないので、浅川(2008)はKS法クラスタ分析という手法を用いて、最適クラスタ数を決定することを提案している。また浅川(2008)は地理情報の可視化にフリーソフト MANDARA を利活用することを指摘している点は本稿の立場と近い。

以上の先行研究では、k-means法が最適クラスタ数を決定しないため、GMM(混合ガウスモデル)やx-means法、KS法クラスタ分析という手法を用いて、最適クラスタ数を決定することが焦点となっている。なお、GMMやx-means法はRにも実装されており、普及しやすい手法と言える。k-means法とGMM、x-means法では、k-means法の方が計算時間は短いという利点がある。既に述べたように、GMM、x-means法ではk-meansより計算時間はかかるが、最適クラスタ数を決定できるという利点がある。

平原(2021)は福岡都市圏のデータを使って、都市社会学の成果を応用してk-meansでクラスタリングを行った研究である。

小阪&村舘(2016)は2010年国勢調査の千葉市データをもとに、GMM(混合ガウスモデル)の手法で小地域のクラスタリングを行っている。GMM(混合ガウスモデル)はベイズ情報量基準を比較することで最適クラスタ数を算出するところがメリットの手法である。

本研究は、小阪&村舘(2016)を受けて2005年、2010年、2015年、2020年の4つの時点の国勢調査のデータで千葉市に関して、GMM(混合ガウスモデル)の手法で小地域のクラスタリングを試みた。

分析にあたってはまず2020年国勢調査のGISデータを入手して、縦軸に小地域(町丁)、横軸に面積、男女別人口総数及び総世帯数、年齢別、男女別人口、世帯別一般世帯数、世帯の家族類型別一般世帯数、住宅の所有の関係別一般世帯数、住宅の建て方別世帯数、産業別及び従業上の地位別就業者数、職業別就業業者数、世帯の経済構成別一般世帯数のデータからなる表を作成した。この表をさらに集計して、変数を取捨選択して、年少人口比率、生産年齢人口比率、高齢化率、第一次産業従事者比率、第二次産業従事者比率、第三次産業従事者比率、ホワイトカラー比率、グレーカラー比率、ブルーカラー比率からなる表を毎年作成した。これをCSVファイルとして読み込んでクラスタ分析を行い、結果を地図上に可視化して解釈を試みる。現時点では2005, 2010, 2015, 2020の4つの国勢調査の時点に関して分析が行うことが可能であり、その結果を同じ千葉市でクラスタの変遷として解釈しようと試みる。また分析がある程度進捗した後で、再度千葉市の行政担当者に分析結果をフィードバックして、議論を重ねたいと希望している。

データの前処理にあたっては、データ数が極端に低い町丁(-)や秘匿地域(X)があることが想定されるが、これらは一括してゼロの値に置き換える。また分析にあたって、極

によって比較し、収束するまで適用するという手法である。

端に人口が少ない地域は白色のクラスタに分類する。

データの前処理をする前の 2020 年の千葉市の小地域のデータは小地域が 519、面積、人口、世帯等属性が 131 項目から構成されている。

各国勢調査から、年少人口比率、生産年齢人口比率、高齢化率、第一次産業従事者比率、第二次産業従事者比率、第三次産業従事者比率、ホワイトカラー比率、グレーカラー比率、ブルーカラー比率を集計し、さらに経度緯度、人口密度を加えた表を作成し、まず 2020 年のデータで分類し、その結果を学習して、2005 年、2010 年、2015 年に適用する⁹。

なお分析、結果の応用が期待される分野の一つが都市計画であり、千葉市の場合、その基礎資料は千葉市(2016)『千葉市都市計画マスタープラン』である。千葉市(2016)によると、千葉市は人口減少・少子超高齢化社会の到来を見通して、集約型都市構造（コンパクトシティ）を目指している。千葉市は、3つの都心部（千葉都心、幕張副都心、蘇我副都心）を設定して、都市機能集積ゾーン、臨海ゾーン、生活・環境調和ゾーン、自然共生ゾーンからなる4つのゾーンに基づき、都市計画を立案している。

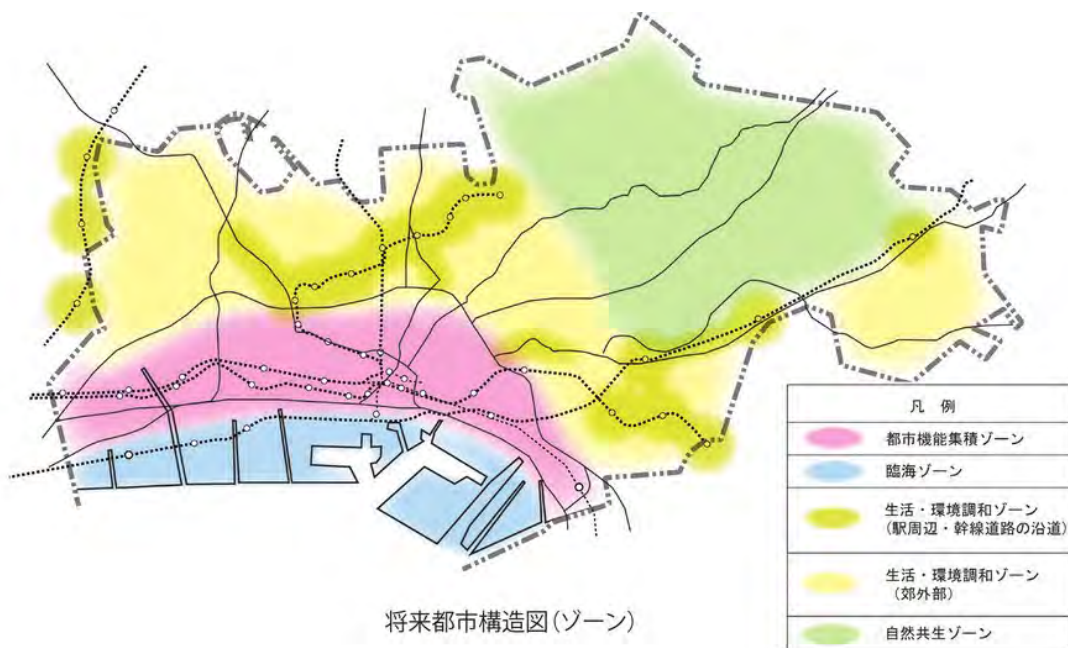


図 3-1 千葉市の将来都市構造図（ゾーン）

出所) 千葉市 (2016) 千葉市都市計画のマスタープラン p. 23 から引用

⁹ 緯度経度は空間的隣接性を表す情報として有用だが、一部データが MANDARA 上で整数倍されて処理されてしまうこと（町丁字等の重複フラグがある町丁は重複した緯度経度が表示される）が発見されたため、正しく修正した緯度経度をクラスタリングの入力値とした。

千葉市の将来都市構造図（ゾーン）によると都市機能集積ゾーン、臨海ゾーン、生活・環境調和ゾーン、自然共生ゾーンの4つのゾーンに千葉市の将来都市構図は分類されるといふ。このうち、都市集積ゾーンは電車の駅が集中しており、逆に自然共生ゾーンは駅から外れたゾーンになっている。また生活環境・調和ゾーンは2つのゾーンから構成されており、駅周辺・幹線道路の沿道と郊外部に分類されている。

4. 分析結果

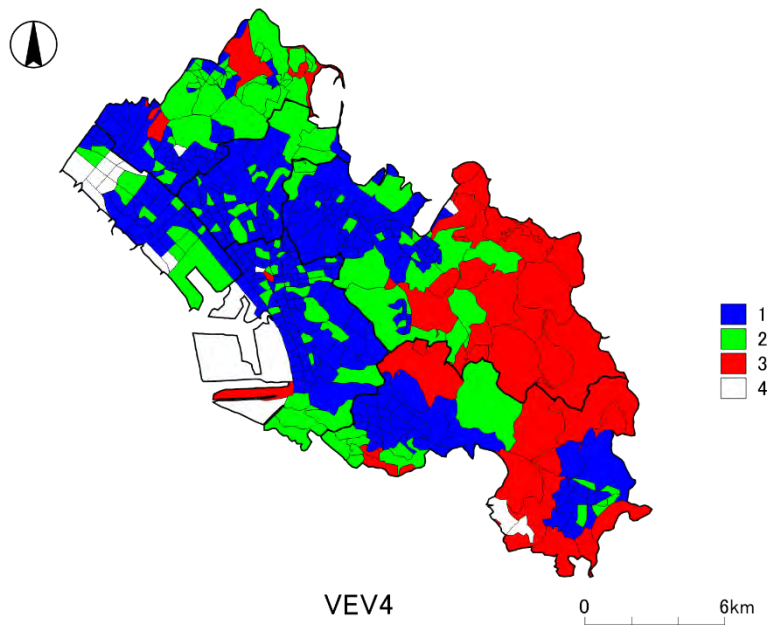


図 4-1. 2020 年の千葉市のクラスタリング結果

表 4-1 2020 年の千葉市の各クラスタの特徴

クラスタ	個数	人口密度	年少人口比率	生産年齢人口比率	老年人口比率	第一次産業従事者人口比率	第二次産業従事者人口比率	第三次産業従事者人口比率	ホワイトカラー人口比率	グレーカラー人口比率	ブルーカラー人口比率	特徴
1	341	0.0093	11.0	61.3	27.7	0.4	15.7	83.9	47.3	28.2	24.5	人口密集・第三次産業・ホワイトカラー
2	119	0.0081	11.4	58.8	29.8	1.0	18.9	80.1	41.3	26.4	32.3	年少人口・第二次産業
3	39	0.0003	6.7	48.6	44.7	15.3	18.0	66.6	27.9	23.4	48.6	高齢化・第一次産業・ブルーカラー
4	20	0.0001	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	人口ゼロ
総計	519	0.0080	10.3	57.4	28.4	1.6	16.0	78.5	42.6	26.3	27.2	

2020 年の千葉市は図 4-1 のように混合ガウスモデル（アルゴリズム名 VEV）で BIC（ベイズ情報量基準）の観点から、4 つのクラスタに分類されたが、人口規模の観点からゼロでは

ない、分析対象となるクラスタは3つである。

クラスタ1（青）は人口密集で、第三次産業比率が高く、ホワイトカラー比率が高い。

クラスタ2（緑）は年少人口比率が比較的高く、第二次産業比率が高い。

クラスタ3（赤）は、高齢化率が高く、第一産業、ブルーカラー比率が高い。

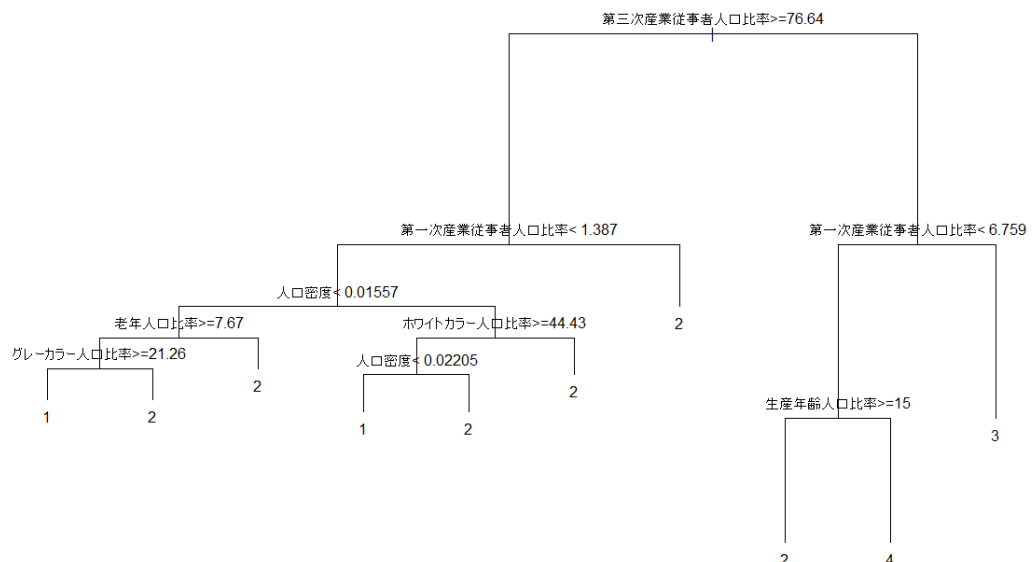


図 4-2. 2020 年の千葉市のクラスタリング結果の決定木による解釈

決定木によるクラスタリング結果の解釈(図 4-2)は、まず第三次産業従事者人口比率 (76.7%) でクラスタ 1, 2 および 2, 3, 4 に分岐すると言える。この内、クラスタ 4 は人口がほぼゼロで臨海部に多いクラスタであり、ほぼ無視できる。つまり第三次産業従事者人口比率で、クラスタ 1 & 2、2 & 3 に分岐していると言える。クラスタ 1 の特徴は、第 3 次産業、つまりサービス産業の比率が多いことにある。一方で、クラスタ 3 の特徴は、第 1 次産業の比率が高く、第三次産業の比率が低い。

次に第一次産業従事者人口比率で分岐している。第三次産業従事者人口比率でクラスタ 1, 2 に分類された内、第一次産業従事者人口比率が (1.4%) 以上だと、クラスタ 2 に分類される。この分岐は、クラスタ 1 とクラスタ 2 の分岐が、産業分類別従事者比率によるこ

とを示唆している。第三次産業従事者人口比率でクラスタ 2 & 3 に分類された内、第一次産業従事者比率が (6.8%) 以上であるとクラスタ 3 に分類される。第一次産業従事者比率が (6.8%) 未満であるとクラスタ 2 に分類される。

ここまででクラスタ 3 を決定づける特徴は人口密度および産業分類別従事者比率であることがわかる。第三次産業従事者人口比率が 76.7% 未満で、第一次産業従事者比率が 6.8% 以上であるとクラスタ 3 に分類される。クラスタ 3 は第三次産業の比率が比較的 low、第一次産業の比率が高い。千葉市に漁業従事者が少ないことを考慮すると、クラスタ 3 は農業従事者の比率が多いクラスタである。

クラスタ 1 と 2 の分岐を特徴づけるものとしては、人口密度、老年人口比率、ホワイトカラー比率、グレーカラー比率がある。クラスタ 1 とクラスタ 2 を比較すると、クラスタ 1 の方がグレーカラー比率に関して高いことがわかる。またホワイトカラー比率もクラスタ 1 の方が高い。

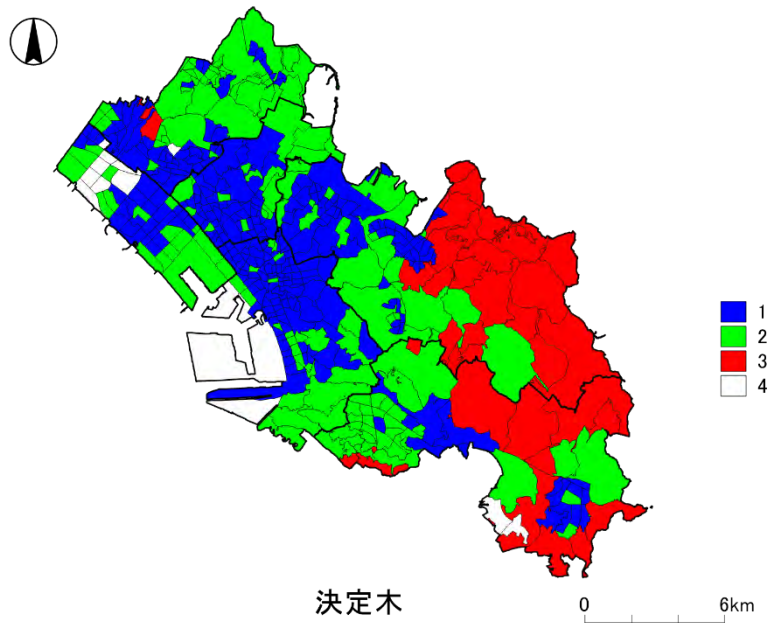


図 5-1. 2005 年の千葉市のクラスタリング結果

2020 年の分類結果を決定木で学習し、2005 年のデータに適用した結果である。

表 5-2 2005 年のクラスタ分類結果

クラスタ	個数	人口密度	年少人口比率	生産年齢人口比率	老年人口比率	第一次産業従事者人口比率	第二次産業従事者人口比率	第三次産業従事者人口比率	ホワイトカラー人口比率	グレーカラー人口比率	ブルーカラー人口比率	特徴
1	289	0.0090	11.5	68.8	19.7	0.2	16.9	82.8	46.6	31.5	21.9	人口密・第三次産業・ホワイトカラー
2	149	0.0077	15.3	69.9	14.9	1.0	22.4	75.3	42.0	28.5	28.2	年少人口・第二次産業
3	35	0.0003	8.9	60.6	30.5	21.5	20.3	58.2	26.1	20.5	53.4	高齢化・第一次産業・ブルーカラー
4	16	0.0000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	人口ゼロ
総計	489	0.0077	12.1	66.3	18.3	2.0	18.3	76.1	42.2	28.8	25.4	

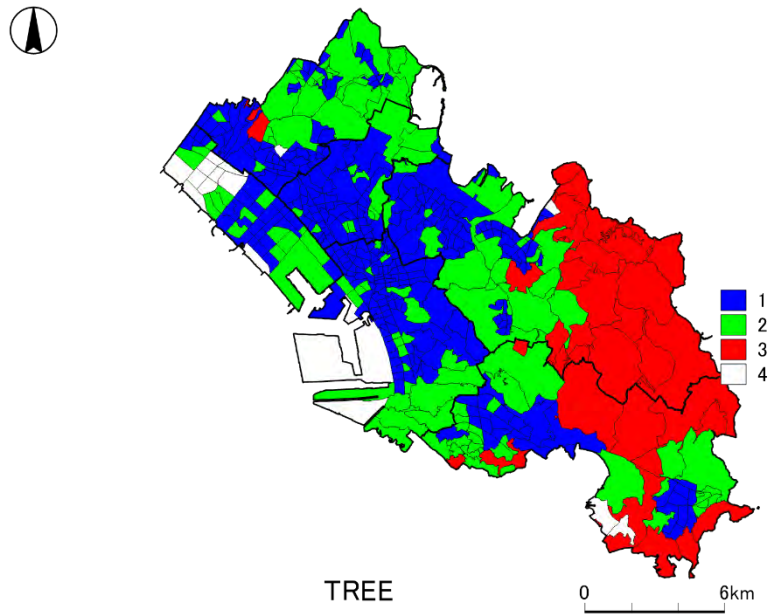


図 6-1. 2010 年の千葉市のクラスタリング結果

図 6-1 は 2010 年のデータを基に、2020 年で作ったモデル（決定木）で分類した結果である。

表 6-2 2010 年のクラスタ分類結果

クラス	個数	人口密度	年少人口比率	生産年齢人口比率	老年人口比率	第一産業従事者人口比率	第二産業従事者人口比率	第三産業従事者人口比率	ホワイトカラー人口比率	グレーカラー人口比率	ブルーカラー人口比率	特徴
1	339	0.0090	11.5	64.8	23.7	0.3	14.9	84.9	43.1	29.5	27.4	人口密・第三次産業・ホワイトカラー
2	122	0.0080	14.6	64.9	20.5	1.1	20.0	78.9	39.4	27.2	33.4	年少人口・第二次産業
3	33	0.0003	7.6	56.8	35.7	16.2	18.2	65.6	24.2	20.8	55.0	高齢化・第一産業・ブルーカラー
4	18	0.0001	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	人口ゼロ
総計	512	0.0079	11.6	62.0	22.9	1.5	15.8	79.2	39.5	27.4	29.6	

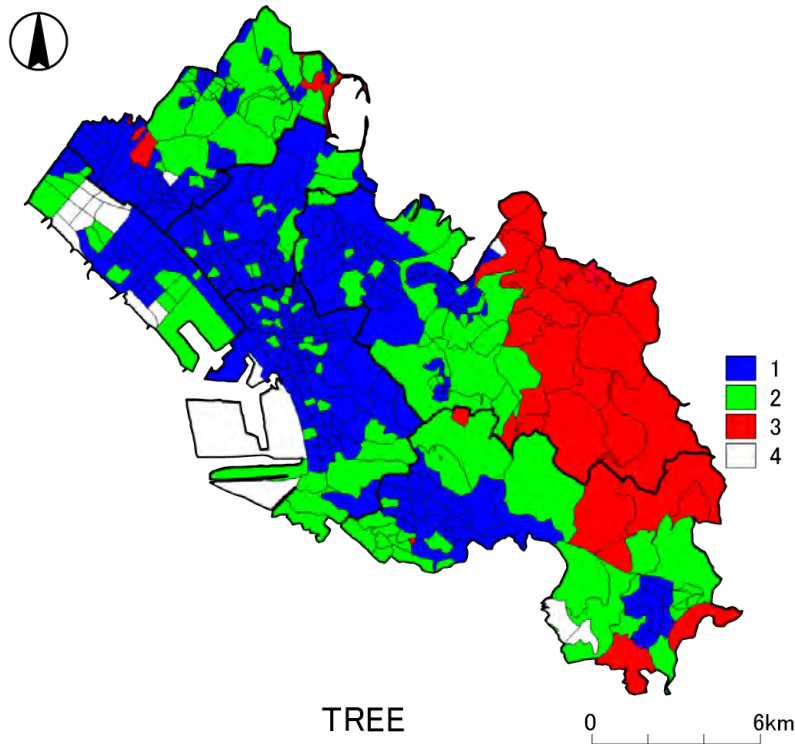


図 7-1. 2015 年の千葉市のクラスタリング結果

図 7-1 は 2015 年のデータを基にした分類結果である。

表 7-2 2015 年のクラスタ分類結果

クラスタ	個数	人口密度	年少人口 比率	生産年齢 人口比率	老年人口 比率	第一次産 業従事者 人口比率	第二次産 業従事者 人口比率	第三次産 業従事者 人口比率	ホワイト カラー人 口比率	グレーカ ラー人口 比率	ブルーカ ラー人口 比率	特徴
1	349	0.0092	11.5	62.6	25.9	0.3	16.3	83.4	44.4	28.7	26.9	人口密・第三次産業・ホワイトカラー
2	120	0.0075	12.4	59.8	27.8	1.4	20.4	78.2	38.8	26.5	34.8	年少人口・第二次産業
3	28	0.0002	6.8	53.0	40.2	19.6	17.6	62.8	23.7	19.5	56.7	高齢化・第一次産業・ブルーカラー
4	18	0.0000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	人口ゼロ
総計	515	0.0080	11.0	59.3	26.2	1.6	16.8	78.2	40.4	26.7	29.4	

5. 考察

クラスタの変遷をまとめた表 8-1 によるとクラスタ 1 (青) はマジョリティで、人口密・第三次産業、ホワイトカラー比率が高い。2005 から 2015 まで増加、2020 に微減している。クラスタ 2 (緑) はセカンドマジョリティで、年少人口比率が高く、第二次産業比率が高い。クラスタ数は一貫して減少している。(少子高齢化)。クラスタ 3 (赤) は高齢化、第一次産業、ブルーカラー比率が高い。2005 年から 2015 年にかけて減少しているものの、2020 年にかけて増加している。

千葉市の都市計画のマスタープランに戻ると、千葉市の都市将来構造図(ゾーン)とのかかわりでは、都市機能集積ゾーンとクラスタ 1 (青) が重なることが多い。また生

活・環境調和ゾーンとクラスタ 2（緑）が重なる。さらに自然共生ゾーンとクラスタ 3（赤）が重なる。臨海ゾーンに関してはクラスタ 4（白）があてはまっている。

クラスタリングの際に特徴量として、人口密度の情報を用いたことで、都市計画におけるゾーン設定と親和的なクラスタリング結果が表れたと言える。

開発の方向性と今回抽出した地区の特性が概ね一致していることやクラスタの面積が変動していることが確認でき、都市計画におけるゾーン設定の際に、小地域クラスタリングが一定の役割を果たせるのではないかと考える。

表 8-1 千葉市の各クラスタに含まれる町丁数の変遷

	(個数)				
クラスタ	2005	2010	2015	2020(基準年)	特徴
1	289	339	349	341	人口密・第三次産業・ホワイトカラー
2	149	122	120	119	年少人口・第二次産業
3	35	33	28	39	高齢化・第一次産業・ブルーカラー
4	16	18	18	20	人口ゼロ
総計	489	512	515	519	

6. 結論

千葉市の町丁に関して、国勢調査のデータから 2020 年に関して混合ガウスモデルで分類したところ、分析の対象となる人口がゼロではないクラスタは 3 つであった。この 2020 年基準のクラスタ分類アルゴリズムを 2005 年、2010 年、2015 年に適用し分析した。4 つの時点に関して分析結果を比較すると、クラスタ 1（青）はマジョリティで、人口密・第三次産業、ホワイトカラー比率が高い。2005 から 2015 まで増加、2020 に微減。

クラスタ 2（緑）はセカンドマジョリティで、年少人口比率が高く、第二次産業比率が高い。クラスタ数は一貫して減少。（少子高齢化）。

クラスタ 3（赤）は高齢化、第一次産業、ブルーカラー比率が高い。2005 年から 2015 年にかけて減少しているものの、2020 年にかけて増加。

このように各クラスタの面積は中長期的には変動、さらには変動の方向が転換点を迎えている可能性がある。東京圏に属する千葉市においても、これまで発展してきたクラスタ 1 の面積が減少し、高齢化や農林業が特徴的なクラスタ 3 が拡大に転じる一方、少子化を反映するクラスタ 2 の減少が続いていることは興味深い。地区ごとの開発方針も影響を受けよう。

今後の課題としては、他の政令市の分析を試みる事が挙げられる。政令市だけではなく、地方の都市や県レベルでの分析に応用することも可能と考える。例えば、防災地図とクラスタリング結果の重ね合わせのように、その他の主題図と重ね合わせて知見を探索することも挙げられる。自治体の政策担当者との対話を通じ、改善を図れば幸いである。

参考文献

- Bishop Christopher, M. (2006). *Pattern recognition and machine learning*. Information science and statistics. New York: Springer.
- Pelleg, D., & Moore, A. W. (2000, June). X-means: Extending k-means with efficient estimation of the number of clusters. In *Icml* (Vol. 1, pp. 727-734).
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal*, 8(1), 289.
- 赤池弘次, 甘利俊一, 北川源四郎, 樺島祥介, 下平英寿, 室田一雄, & 土谷隆. (2007). *赤池情報量規準 AIC—モデリング・予測・知識発見—*. 共立出版.
- 浅川達人. (2008). 社会地区分析再考: KS 法クラスター分析による 2 大都市圏の構造比較. *社会学評論*, 59(2), 299-315.
- 甘利俊一. (2014). *情報幾何学の新展開*. サイエンス社.
- 小阪勇氣, & 村舘靖之. (2016, August). 国勢調査データを用いた小地域クラスタリング. In *IEICE Conferences Archives*. The Institute of Electronics, Information and Communication Engineers.
- 桐村喬. (2007). 小地域の地理的クラスタリング—外れ値処理と空間的スムージング—. *GIS-理論と応用*, 15(2), 81-92.
- 佐藤仁志. (2011). 非階層的クラスタリングによる東京大都市圏の考察. *麗澤経済研究*, 19(1), 123-135.
- 千葉市. (2016). 千葉市都市計画マスタープラン.
(https://www.city.chiba.jp/toshi/seisaku/2016_masterplan.html)
最終訪問 2023/04/18.
- 李燕. (2009). 空間分析におけるクラスター分析と自己組織化マップ手法の比較. *土木計画学研究・論文集*, 26, 77-84.
- 平原幸輝. (2021). 福岡都市圏の空間構造とその変容—市区町村・小地域データを用いたクラスター分析を踏まえて—. *都市政策研究*, (22), 11-20.
- 村舘靖之. (2022). 川崎市の小地域クラスタリング. *第 84 回情報処理学会全国大会講演論文集*, 2022(1), 475-476.

付録 分析に用いる R のコード

- `library("mclust")`
- `r<-read.csv("Chiba.csv",header=TRUE)`
- `t=scale(r)`
- `plot(mclustBIC(t))`
- `mhc <- hc(modelName="VII",data=t)`
- `cl <-hclass(mhc,4)`
- `write.csv(cl,"ccluster.csv")`
- `library(rpart)`
- `tt=merge(r,cl)`
- `tree1<-rpart(cl~.,method="class",data=tt)`
- `plot(tree1) ; text(tree1)`
- `round(predict(tree1,data=r2))`

まず `mclust` のパッケージを読み込む。

次に分析対象となる `csv` ファイルを読み込む。

データを正規化する。

BIC（ベイズ情報量基準）をプロットする。

BIC の観点で適切（BIC が最大）とされたモデル（ここでは VII）でモデルを作る。

クラスタ数は BIC をプロットした時にわかるので、それを元にモデルを分類する。

分類結果を `csv` ファイルで書き出す。

次に `rpart` のパッケージを読み込む。

決定木を計算する。

決定木をプロットする。

テキストをオーバーラップする。

別の年のデータ `r2` を読み込んで、決定木 1 を使い、所属クラスタを予測する。