

川崎市の小地域クラスタリング

村舘 靖之[†]

内閣府 経済社会総合研究所[†]

1. はじめに

統計情報を地図上に可視化し分析する手法は、行政においてもある程度は共有されているようになってきているが、まだ利活用が十分に進んでいるとは言えない。国勢調査のようなまち、ひと、しごとに関する基本的な統計情報を地図上に可視化する場合、データからクラスタ分類を行い、付加価値をつけて可視化する手法を論ずる。川崎市の町丁（小地域）について機械学習の一手法である混合ガウスモデルを用いてクラスタリングを行う。混合ガウスモデルは、クラスタ数決め打ちの k -平均法と異なり、ベイズ情報量基準を参照して、最適なクラスタ数を決定する。本稿の手法は、データ（情報）の圧縮手法であるといえる。本来町丁数×国勢調査の項目数だけあるデータをクラスタに圧縮することを行った。

今回の分析手法や解釈の方法をできるだけ標準化することで、自治体（政令市）の地域カルテ作りに必要なマニュアル作成に本研究が発展する可能性を望みたい。

2. データおよび手法

2-1. データ

まず国勢調査の GIS データセットで境界情報と統計情報（2015 年版）を入手する。本稿では対象を川崎市に設定している。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。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。

2-2. 手法

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数据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 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。GMM では BIC（ベイズ情報量基準）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。

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。

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）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。本稿は、地方自治体の政策立案にフォーカスして、国勢調査を GMM 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ため、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も本稿の立場に近い。

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 法の方が計算時間が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。

小阪&村館(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&村館(2016)を受けて 2015 年国勢調査のデータで川崎市の小地域のクラスタリングを試みた。

3. 適用結果

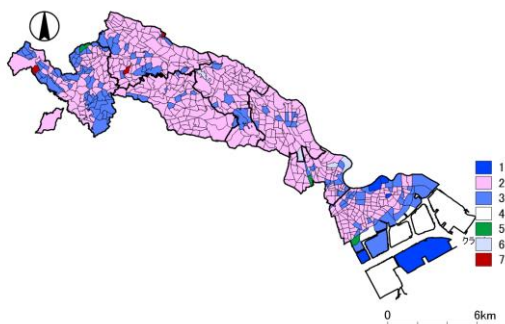


図 1. 川崎市のクラスタリング結果

単純にクラスタリングを国勢調査のデータを投入して行くと人口規模に結果が引っ張られるため、面積（を元に人口密度）、緯度経度、既存行政区を変数に追加し、国勢調査の各変数は人口一人当たり、世帯当たりで総数を除いて行った。

次に分類結果を解釈するため、決定木で分析を行った。

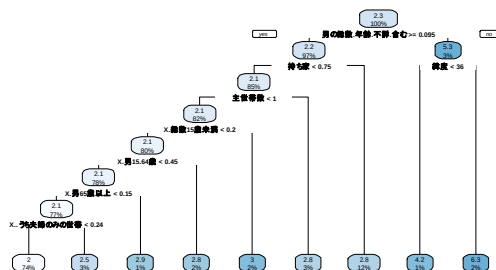


図 2. 決定木による分類パターンの分析

4. 考察

川崎市は BIC の観点から最適クラスタ数を検討するとク

ラスタ数が 1 の結果を除くと、次善の結果はクラスタ数が 7 となった。

川崎市は無中心ないし複数の中心があるという特徴がある。人口規模を排除したため、男女比が特徴量として強調されている。次は緯度、持ち家、子供の比率、生産年齢人口（男性）、男性の高齢者人口比率、夫婦のみの世帯に分岐が発生している。

5. 結論

川崎市は BIC の観点から最適クラスタ数を検討するとクラスタ数が 1 の結果を除くと、次善の結果はクラスタ数が 7 となった。今回の分析では、人口密度および緯度経度、外れ値ダミーを変数に追加している。

本稿の手法は、ある意味データ（情報）の圧縮手法であるといえる。本来町丁数×国勢調査の項目数だけあるデータをクラスタに圧縮することを行った。

一定以上の人口規模が期待される政令市に関しては、今回のような分析は有効なのではないだろうか。規模の小さな自治体だと、町丁の数が限られるので、機械学習に掛けてクラスタリングを行うことは、含意が少なくなってしまうおそれがある。

また分析手法や解釈の方法をできるだけ標準化することで、自治体（政令市）の地域カルテ作りに必要なマニュアル作成に本研究が発展する可能性を望みたく、今後も継続して地域政策立案のための基礎的研究を継続したい。

謝辞

本稿は研究者としての個人の意見であり、所属機関を代表するものではありません。

参考文献

- [1] Bishop Christopher, M. (2006). *Pattern recognition and machine learning*. Information science and statistics. New York: Springer.
- [2] 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).
- [3] 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.
- [4] 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.
- [5] 浅川達人. (2008). 社会地区分析再考: KS 法クラスタ分析による 2 大都市圏の構造比較. *社会学評論*, 59(2), 299-315.
- [6] 甘利俊一. (2014). *情報幾何学の新展開*. サイエンス社.
- [7] 小阪勇氣, & 村館靖之. (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.
- [8] 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.
- [9] 川崎市. (2016). 川崎市総合計画について. (<https://www.city.kawasaki.jp/170/page/0000075895.html>) 最終訪問 2021/08/05.
- [10] 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.
- [11] 千葉市. (2016). 千葉市都市計画マスタープラン. (https://www.city.chiba.jp/toshi/somu/documents/honpen_all.pdf) 最終訪問 2021/08/05.
- [12] 李燕. (2009). 空間分析におけるクラスタ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.