

社会経済データに基づく地域分類の比較検証

貞広幸雄

Variation in classification of regions

Yukio Sadahiro

Abstract: The region is an essential concept in GIS and geography. Regional classification has been conducted over a long period based on a wide variety of regional properties. Recent development of data acquisition techniques and improvement of availability in spatial data permit us to consider a wider variety of variables from physical to social, economic, and cultural properties. This paper briefly discusses the similarity and difference among regional classification based on different variables. Similar classification results implies that their schemes are complementary with each other so that spatial data of low accessibility can be substituted with those of high accessibility in terms of regional classification.

Keywords: 地域分類 (regional classification), 空間データ (spatial data), 社会経済データ (socio-economic data)

1. はじめに

地域分類は、地理学における基本的な概念の一つである。同質地域の抽出のために、定性的、定量的、主観的、客観的、極めて多数の分類手法が開発されてきている。いずれの手法も、依拠する地域特性やデータが異なり、分類結果は極めて多様である。

近年急速に普及しつつある社会経済データは、こうした地域分類のための新たな手段である。公的データのみならず、消費支出推計や自動車保有台数、貯蓄推計など、地域特性を知るための様々なデータが、比較的容易に入手できるようになってきている。さらには、Mosaic Japan (アクトン・ワインズ (株)) や Cameo (日本 GMAP (株)) など、こうしたデータに基づいて、地域を分類したジオデモグラフィック

クデータも販売されるようになり、マーケティング分野などを中心に、幅広く用いられてきている。

本稿では、このように急速に普及した社会経済データに基づいて、様々な地域分類を実証的に実施し、研究分野での有用性を検討することを目的とする。最も代表的な社会経済データである国勢調査を始めとし、官民間わず、種々のデータに基づいた地域分類を行い、その結果を比較検討する。

2. 地域分類

本稿では、平成 17 年度国勢調査小地域統計に含まれる千葉県の全町丁目 5696 地域を対象とし、国勢調査を始めとする各種空間データによる地域分類を行う。分類には、国勢調査小地域統計、千葉県都市計画基礎調査、Mosaic Japan、消費支出推計データ ((株) JPS) を利用している。地域分類は表-1 に示す変数に基づき、15 通りの方法で実施した。

貞広幸雄 〒113-8656 東京都文京区本郷 7-3-1

東京大学大学院工学系研究科都市工学専攻

Phone: 03-5841-6273

E-mail: sada@ua.t.u-tokyo.ac.jp

分類方法には Ward 法と K-means 法を用い, 分類数は Mosaic Japan の大分類+未分類の, 合計 12 とした.

表- 1 地域分類法

分類基準	分類の依拠する変数	変数の数
Mosaic	国勢調査, 推計年収等	-
年齢構成	10歳階級別人口構成比	8
就業産業構造	産業別人口構成比	19
居住年数	居住年数別世帯構成比	4
住宅所有形態	住宅所有形態別世帯構成比	4
土地利用比率	土地利用構成比	-
都市地域分類	土地利用, 建築年数等	-
(以下は全て, 消費支出推計データにおける支出額構成比)		
消費実態	家計における主たる支出	10
交通手段	交通手段	3
蛋白源	蛋白源となる食品	3
肉類	肉類	3
魚介類	魚介類	15
果物類	果物類	9
調味料	調味料	5
酒類	酒類	4

なお, Ward 法と K-means 法による分類結果は, いずれの分類基準においても大変類似していたことから, 以下では Ward 法による結果のみを示す. 参考のために, いくつかの分類結果を文末に示す.

3. 分類結果の比較

次に, 分類結果の定量的比較を行う. いま, 2 つの分類法 C_A と C_B を考える. 全ての地域の対について, C_A と C_B のいずれの方法でも同じ分類に属するもの, C_A では同じ分類に属するが, C_B では異なる分類に属するもの, C_A では異なる分類に属するが, C_B では同じ分類に属するもの, C_A と C_B のいずれの方法でも異なる分類に属するもの, それぞれの個数を v_{11} , v_{10} , v_{01} , v_{00} とする.

分類結果比較のための代表的な指標として, 修正 Rand 指標と, 修正 Wallace 指標がある. それぞれ, 以下の式で定義される.

$$R = \frac{v_{11} + v_{00} - E[R_0]}{v_{11} + v_{10} + v_{01} + v_{00} - E[R_0]}$$

$$W_{AB} = \frac{\frac{v_{11}}{v_{11} + v_{10}} - E[W_{AB0}]}{1 - E[W_{AB0}]}$$

$$W_{BA} = \frac{\frac{v_{11}}{v_{11} + v_{01}} - E[W_{BA0}]}{1 - E[W_{BA0}]}$$

但し $E[R_0]$, $E[W_{AB0}]$, $E[W_{BA0}]$ は, それぞれランダムな分類割り当てにおける, Rand 指標と Wallace 指標の期待値である (詳しくは Wagner and Wagner (2007) 及び Severiano *et al.* (2011) を参照のこと). いずれも, 0~1 の変域を持ち, 大きな値ほど分類結果が類似していることを表す.

以下, 計算結果を表-2 に示す. この表より, まず, 消費支出推計データに基づく分類は, いずれも類似していることが分かる. 他方, 他の社会経済データに基づく分類は, いずれも大きく異なっており, 利用する変数に分類結果が強く依存することを示唆している. 土地利用比率と都市地域分類は, いずれも物理的環境に基づく分類であることから, 一定程度, 分類結果に類似性が認められる.

次に, 得られた指標値を類似度行列と見なし, 分類手法のクラスタリングを行う. いずれの指標値を用いても結果が類似していたことから, ここでは修正 Wallace 指標に基づく結果のみを図-1 に示す.

この図より, まず, 消費支出推計データに基づく分類の類似性の高さを確認することができる. 他方, 就業産業構造に基づく分類が他と大きくかけ離れていることが分かるが, これは, 図-2 においても同様である. 特に千葉県の半島部においてその乖離が顕著であるが, これらは主として人口の非常に少ない地域であり, 他の分類方法と比べて, 変数の数が大きいことが原因ではないかと考えられる.

表-2 分類結果の類似性. (a) 修正 Rand 指標, (b) 修正 Wallace 指標

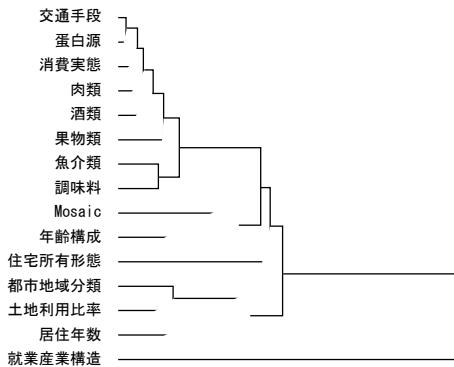
	Mosaic	年齢	就業	居住	住宅	土地	都市	消費	交通	蛋白	肉	魚介	果物	調味	酒	
Mosaic		.146	.034	.029	.058	.033	.078	.092	.091	.099	.094	.117	.101	.117	.093	
年齢構成		.146		.022	.023	.019	.022	.023	.048	.047	.043	.048	.056	.054	.050	.048
就業産業構造		.034	.022		.008	.000	.020	.000	.000	.000	.000	.000	.002	.000	.000	.000
居住年数		.029	.023	.008		.021	.072	.091	.008	.009	.009	.011	.008	.010	.008	
住宅所有形態		.058	.019	.000	.021		.013	.074	.028	.028	.031	.028	.035	.029	.051	.026
土地利用比率		.033	.022	.020	.072	.013		.273	.010	.010	.010	.010	.014	.011	.012	.010
都市地域分類		.078	.023	.000	.091	.074	.273		.028	.028	.032	.029	.038	.029	.049	.028
消費実態		.092	.048	.000	.008	.028	.010	.028		.701	.693	.636	.436	.510	.360	.654
交通手段		.091	.047	.000	.009	.028	.010	.028	.701		.752	.641	.453	.493	.314	.560
蛋白源		.099	.043	.000	.009	.031	.010	.032	.693	.752		.617	.411	.560	.326	.587
肉類		.094	.048	.000	.008	.028	.010	.029	.636	.641	.617		.439	.523	.339	.552
魚介類		.117	.056	.002	.011	.035	.014	.038	.436	.453	.411	.439		.437	.459	.422
果物類		.101	.054	.000	.008	.029	.011	.029	.510	.493	.560	.523	.437		.379	.494
調味料		.117	.05	.000	.010	.051	.012	.049	.360	.314	.326	.339	.459	.379		.357
酒類		.093	.048	.000	.008	.026	.010	.028	.654	.560	.587	.552	.422	.494		.357

(a)

	Mosaic	年齢	就業	居住	住宅	土地	都市	消費	交通	蛋白	肉	魚介	果物	調味	酒	
Mosaic		.138	.046	.027	.075	.036	.116	.088	.089	.095	.096	.143	.111	.179	.090	
年齢構成		.156		.033	.022	.027	.025	.037	.048	.049	.044	.052	.074	.064	.083	.049
就業産業構造		.027	.017		.006	.000	.017	.000	.000	.000	.000	.002	.000	.000	.000	
居住年数		.031	.023	.013		.030	.084	.150	.008	.009	.009	.009	.015	.010	.017	.009
住宅所有形態		.047	.015	.000	.016		.011	.083	.022	.022	.024	.023	.033	.026	.058	.021
土地利用比率		.031	.019	.025	.063	.015		.370	.009	.009	.009	.010	.016	.012	.016	.009
都市地域分類		.058	.017	.000	.066	.067	.217		.020	.021	.024	.022	.033	.023	.050	.020
消費実態		.096	.047	.000	.008	.039	.011	.043		.709	.697	.677	.561	.587	.582	.659
交通手段		.094	.046	.000	.009	.037	.011	.043	.693		.749	.674	.574	.561	.500	.558
蛋白源		.102	.042	.000	.009	.042	.011	.051	.689	.756		.653	.524	.640	.523	.588
肉類		.092	.044	.000	.007	.035	.011	.041	.600	.610	.585		.525	.564	.506	.524
魚介類		.099	.045	.002	.009	.037	.013	.045	.357	.374	.338	.378		.400	.557	.347
果物類		.093	.047	.000	.007	.034	.011	.038	.451	.440	.497	.488	.482		.515	.440
調味料		.087	.036	.000	.007	.045	.009	.048	.260	.229	.237	.255	.391	.300		.260
酒類		.096	.047	.000	.008	.035	.011	.043	.650	.563	.586	.583	.537	.565	.572	

(b)

図-1 分類手法の類似性を示すデンドログラム



次に、分類数を 12 から 4 にした場合の比較を行う。紙面の都合上、ここでは修正 Rand 指標に基づく評価だけを掲載する。

表-3 分類数 4 の場合における修正 Rand 指標

	年齢構成	就業産業構造	居住年数	住宅所有形態	消費実態	交通手段	蛋白源	肉	魚介類	果物類	調味料	酒類
年齢構成	.047	.047	.042	.098	.096	.122	.118	.097	.128	.118		
就業産業構造	.047		.026	.019	.015	.012	.011	.024	.010	.019	.018	
居住年数	.047		.026		.043		.016	.012	.012	.012	.002	
住宅所有形態	.042		.019		.043			.035	.025	.021	.033	.042
消費実態	.098		.015		.016			.035		.486	.582	.380
交通手段	.078		.012		.025			.486		.583	.242	.444
蛋白源	.096		.011		.021			.582		.255	.292	.807
肉	.122		.024		.001			.380		.242	.271	.679
魚介類	.118		.024		.012			.441		.292	.298	.410
果物類	.097		.010		.012			.568		.386	.328	.328
調味料	.128		.019		.002			.405		.291	.320	.880
酒類	.118		.018		.000			.412		.207	.299	.683

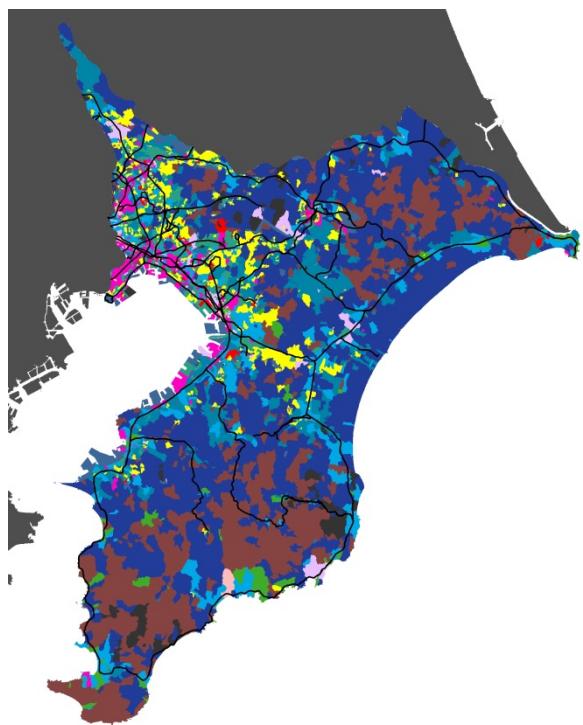
表-3 を見る限り、全体的な傾向は分類数 12 の場合と同様である。消費支出推計データに基づく分類結果が互いに類似している一方、他の社会経済データに基づく分類は、いずれも大きく異なっている。しかしながら、個々の数値を見ると、全体的にやや大きめの値になっており、且つ、値の分散も拡大している。これは、各グループのサイズが大きくなるに伴って、分類が不安定になることに起因するのではないかと考えられる。

4. おわりに

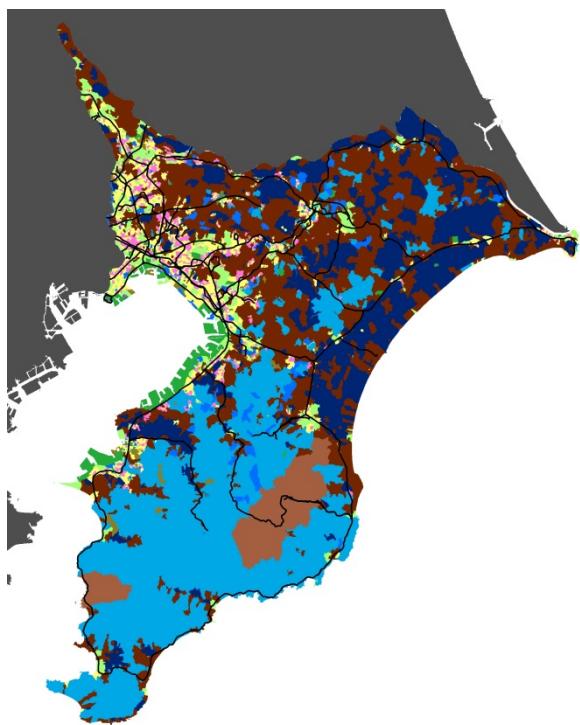
本稿では、様々な社会経済データに基づく地域分類を比較検討した。消費支出推計データという、比較的類似したデータの範疇であれば、分類結果は安定する場合もあるが、異なる種類のデータの場合、分類結果は大きく異なることが多い。地域分類の目的に応じた、適切なデータ選択が肝要であると言える。

参考文献

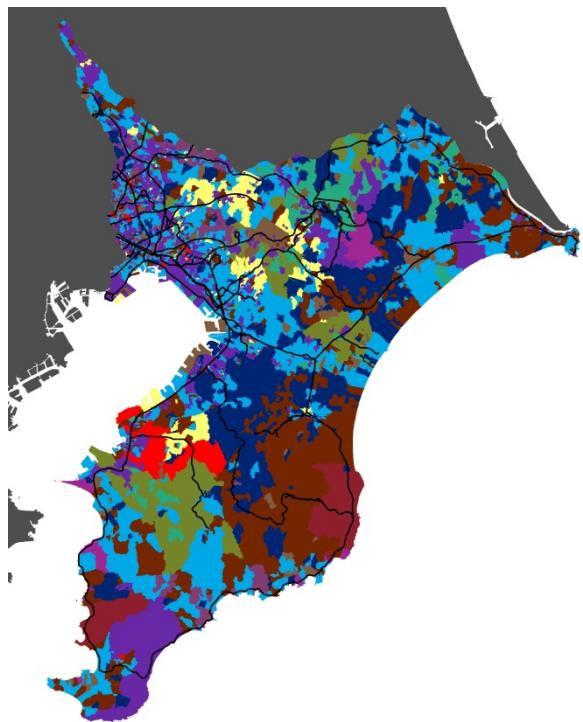
- Wagner S and Wagner D 2007 Comparing clusterings – An overview *Universität Karlsruhe Technical Report 2006-04*
 Severiano A, Pinto FR, Ramirez M, and Carriço J 2011 Adjusted Wallace as a measure of congruence between typing methods *Journal of Clinical Microbiology* **49** (11) 3997-4000



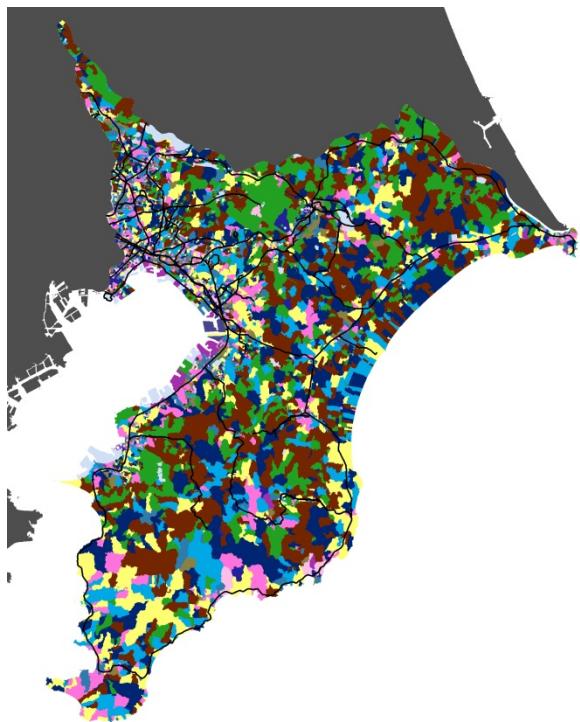
(a)



(b)



(c)



(d)

図-2 (a) Mosaic Japan, (b) 就業産業構造, (c) 土地利用比率, (d) 蛋白源となる食品の支出構成比に基づく分類結果