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）
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。

<table>
<thead>
<tr>
<th>分類基準</th>
<th>分類の基準する変数</th>
<th>変数の数</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mosaic</td>
<td>国勢調査、推計年収等</td>
<td>-</td>
</tr>
<tr>
<td>年齢構造</td>
<td>10歳階級別人口構成比</td>
<td>8</td>
</tr>
<tr>
<td>就業産業構造</td>
<td>産業別人口構成比</td>
<td>19</td>
</tr>
<tr>
<td>居住年数</td>
<td>居住年数別世帯構成比</td>
<td>4</td>
</tr>
<tr>
<td>住宅所有形態</td>
<td>住宅所有形態別世帯構成比</td>
<td>4</td>
</tr>
<tr>
<td>土地利用比率</td>
<td>土地利用構成比</td>
<td>-</td>
</tr>
<tr>
<td>都市地域分類</td>
<td>土地利用、建築年数等</td>
<td>-</td>
</tr>
</tbody>
</table>

（以下は全て、消費支出推計データにおける支出額構成比）

<table>
<thead>
<tr>
<th>消費実態</th>
<th>家計における主たる支出</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>交通手段</td>
<td>交通手段</td>
<td>3</td>
</tr>
<tr>
<td>蛋白源</td>
<td>蛋白源となる食品</td>
<td>3</td>
</tr>
<tr>
<td>肉類</td>
<td>肉類</td>
<td>3</td>
</tr>
<tr>
<td>魚介類</td>
<td>魚介類</td>
<td>15</td>
</tr>
<tr>
<td>果物類</td>
<td>果物類</td>
<td>9</td>
</tr>
<tr>
<td>調味料</td>
<td>調味料</td>
<td>5</td>
</tr>
<tr>
<td>酒類</td>
<td>酒類</td>
<td>4</td>
</tr>
</tbody>
</table>

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用いるために、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。

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

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

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

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。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 消費 交通 酒 狭 食物業 調味 洋

Mosaic .146 .034 .029 .058 .033 .078 .092 .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 .010

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 .014 .011 .012 .010 .010

都市地域分類 .078 .023 .000 .091 .074 .073 .028 .028 .032 .029 .038 .029 .049 .028

消費実態 .002 .048 .000 .008 .028 .010 .028 .701 .693 .636 .436 .510 .360 .545

交通手段 .091 .047 .000 .009 .028 .010 .028 .701 .752 .641 .453 .493 .314 .560

年齢 .009 .043 .000 .009 .031 .010 .032 .693 .752 .617 .411 .560 .326 .587

肉類 .068 .048 .000 .008 .028 .010 .029 .836 .641 .617 .429 .323 .339 .552

果物類 .117 .056 .002 .011 .035 .014 .038 .436 .453 .411 .439 .431 .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 .101 .051 .129 .049 .360 .314 .326 .339 .459 .379 .357

酒類 .093 .048 .000 .008 .026 .010 .023 .504 .506 .587 .552 .422 .494 .357

(a)

Mosaic 年齢 就業 居住 住宅 土地 消費 交通 酒 狭 食物業 調味 洋

Mosaic .138 .046 .027 .075 .036 .116 .008 .089 .095 .096 .143 .111 .179 .090

年齢構成 .146 .022 .023 .019 .022 .023 .048 .047 .043 .048 .056 .054 .050 .048

就業産業構造 .027 .017 .006 .000 .017 .000 .000 .000 .000 .000 .000 .000 .000 .000

居住年数 .031 .023 .013 .030 .046 .015 .008 .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 .016

都市地域分類 .058 .017 .000 .046 .061 .217 .020 .021 .024 .022 .033 .023 .059 .020

消費実態 .096 .047 .000 .008 .039 .011 .043 .709 .697 .677 .561 .587 .582 .589

交通手段 .094 .046 .000 .009 .037 .011 .043 .893 .749 .674 .574 .561 .500 .558

年齢 .102 .042 .000 .009 .042 .011 .051 .589 .756 .653 .524 .540 .523 .588

肉類 .092 .044 .000 .007 .035 .011 .041 .600 .610 .585 .525 .564 .506 .524

果物類 .099 .045 .002 .009 .037 .013 .045 .307 .314 .338 .378 .400 .557 .547

果物類 .093 .047 .000 .007 .034 .010 .038 .471 .440 .507 .488 .482 .515 .440

調味料 .087 .036 .000 .007 .045 .009 .048 .280 .292 .237 .255 .391 .300 .260

酒類 .096 .047 .000 .008 .035 .011 .043 .500 .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

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


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
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