

対応解析とクラスタリングとの統合解析システムによる政治分析の試み¹

辻 光宏

関西大学総合情報学部

クロス表形式データに対して、コレスポネンス解析（対応解析）は、各項目の間の距離をカイ二乗距離で測定することにより、低次元に縦列と横列の変数を同時に表示し、全体的な概観を示すことができる。それに対応した凝集型クラスタリングの Ward 法を適用することによって、そこに内在する階層的な構造を同じ視点から調べることができる。

この報告では、得られた結果を総合してビジュアルに表示することにより、よりデータ構造を理解しやすくすることを目指している。

扱うデータは、政治学での投票行動に関する投票データであり、解析の指針として、地域と政党との関連を総合的に評価することを目指している。

Exploring the Political Field by the Integrated Statistical System with Correspondence Analysis and Clustering

Mitsuhiro Tsuji

Kansai University

The complementarity between correspondence analysis and clustering in the statistics will be applied to the cross-table data. Our purpose is to visualize the results in order to get the basic comprehension of the data structure. We shall analyze the results of some elections for the National Diet of Japan.

1. クロス表データに関するオーディネーション手法

多変量データを視覚によって吟味することを目指したオーディネーション手法の代表的なものは主成分解析である。これは、 n 個の変数と p 個の標本から構成される $n \times p$ の数値データ行列 R を対象としている。この n 個の変数と p 個の標本を低次元にプロットする解析手法としてバイプロットも報告されている。

一方、 n 個のカテゴリー項目と p 個のカテゴリー項目から構成される $n \times p$ の頻度データ行列を、クロス表データ行列 K と考える。例えば、後述の投票データを考えると、 n 個の政党に対する p 個の選挙区での得票数（頻度）のデータ行列が、これにあたる。

2. コレスポネンス解析とクラスタリング

クロス表データ行列をコレスポネンス解析で評価することによって、カテゴリー項目とカテ

¹ 本研究は、2000 年度関西大学在外研究（the Department of Mathematics, University of Essex）の成果に基づいている

ゴリー項目との関連について相対的な位置づけを概観することができる。すなわち、カテゴリーそれぞれの分布（プロファイル）をできる限り保存するように、高次元空間上のポイントとしてのクロス表データの行と列を、次式の Pearson のカイ二乗統計量（横列方向が $d^2(i, i')$ 、縦列方向が $d^2(j, j')$ ）で表現し、解釈を容易にするためにベクトル分解により低次元に射影する。

$$d^2(i, i') = \sum_{j=1}^p \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{.i}} - \frac{f_{i'j}}{f_{.i'}} \right)^2, \quad d^2(j, j') = \sum_{i=1}^n \frac{1}{f_{.i}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2$$

ここで、 $f_{ij} = k_{ij}/k$ 、 $f_{.i} = \sum_{j=1}^p f_{ij}$ 、 $f_{.j} = \sum_{i=1}^n f_{ij}$ である。

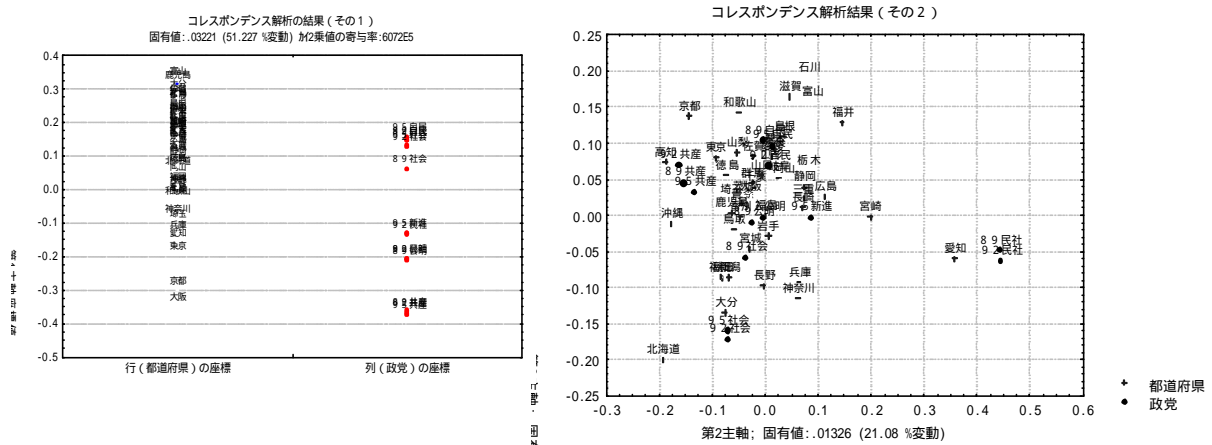
一方、Ward 法による凝集型クラスター解析を行なうことによって、階層的なグループを解明することができ、カテゴリー項目の間の局所的な構造を明確にするとともに、はずれ値を検出することもできる (Lebert, 1994)。これらの両手法によるそれぞれの情報圧縮過程では同じ固有値が検出されることも知られている。

また、コレスポネンス解析によって低次元で表示された項目値プロット図にさらに、クラスター解析結果の樹状図を付記することにより、全般的な関連と部分的な関連を明瞭にすることができる (Greenacre, 1988)。

3. 政治分析への応用

選挙に関する日本での投票データの代表的な参議院の比例代表データを取り上げる。ここでは、1989年、1992年、1995年の各都道府県でのデータを対象にする。政党は、自民党、社会党、公明党、共産党、民社党とする。

コレスポネンス解析で投票データを評価することによって、政党と選挙区の関連について相対的な位置づけを概観することができる (Fehlen, 1998)。コレスポネンス解析の結果を低次元で表現する際に獲得した各次元の主軸に対する意味づけを行なうことにより、各政党および各都道府県の特徴についての解釈を助けることができる。この意図のもとに、各都道府県が表現される主軸座標データと相関の高いデータを探索した。各都道府県の特長を表す変数として、主成分解析でとり上げられた地域特性データ (小林, 1985) を採用した。



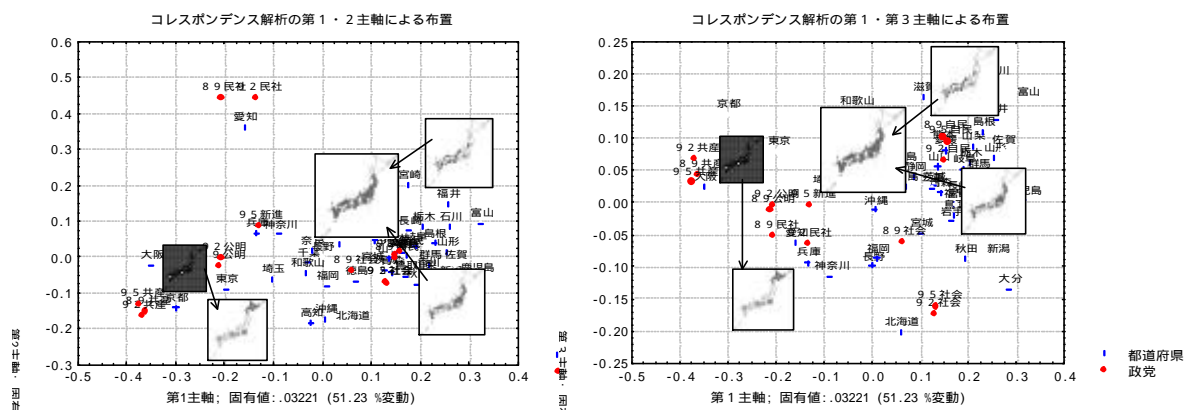
コレスポネンス解析の結果として、第1主軸に対する都道府県と政党の布置を検証すると、主成分解析などで従来から得られている都市化の要因が表示されている。すなわち、第1主軸のマイナス方向に大都市圏のポイント（愛知・東京・京都・大阪、共産党・公明党・民社党）が並んでおり、当時の2大政党（自民・社会）がプラス方向に位置している。

そのため、ここでは、第2主軸と第3主軸による布置が注目される。すなわち、第2主軸（第2次産業就業率と相関が高い）によって大都市に強い少数野党が分かれている。また、第3主軸（前述の地域特性変数との相関はほとんど無い）によって、自民党と社会党が別方向に布置されている。クラスター解析の結果と照合して、これらの布置の検討を進めることによって、新たな構造を把握することができる。

4. 統合による統計グラフィックスの試み

コレスポネンス解析を行うことによって、各項目をプロットする低次元上の相対的な座標値を定めることができる。が、局所的なグループ融合を示すことのできるWardの凝集型クラスタリング結果を、いかにしてコレスポネンス解析結果と統合することができるかを、いくつか試行した。

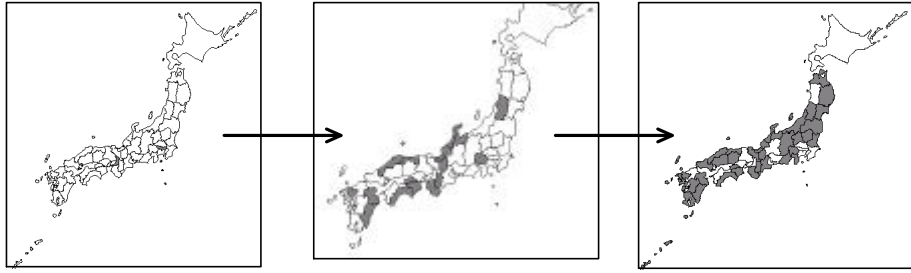
代表的なクラスターを地図で表現し、地域的な要因も考えに入れながら全体的なクラスタリング過程を洞察することが可能となった。ただし、地図そのものが非常に小さいものとなったので、地図上の微妙な変化を捉えることができない危険性を含んでいる。



凝集型クラスタリングの過程は、以下のように説明することができる。

- (1) 初期状態では、1つの変数が1つのクラスターを構成している。
- (2) 最も近いクラスター同士を融合する。
- (3) 融合されたクラスターとその他のクラスターとの間の距離を何らかの基準で求める。Ward法の場合には、融合でのクラスター内分散を最小にするように計算する。
- (4) 全体が1つのクラスターになるまで、(2)(3)の過程を繰り返す。

この(2)(3)の過程を明確に表すことができるのが、クラスター解析でよく用いられる樹状図である。しかし、この報告では、融合過程を時間の推移と捉えて、動画コンテンツでクラスター融合の課程を表現することを試行した。その結果は、研究会で紹介する。



5 . 今後の課題

クロス表データは、アンケート調査などで頻繁に現れるデータである。この研究の良否を判断するには、いろいろな分野のクロス表データを取り扱っていく必要がある。

政治学の分野では、実証的研究が中心であるために、ポピュラーな統計解析手法（例えば主成分解析）を用いる傾向が強い。コレスポネンス解析を中心としたこの一連の手法の展開の有効性を検証するためには、都道府県内の市区町村レベルでの解析を試みるとともに、最近の選挙傾向、例えば無党派層、を意識した分析にも取り組む必要がある。

6 . 参考文献

- [1] 小林良彰(1985) . 計量政治学 . 成文堂 .
- [2] 三宅一郎(1995) . 日本の政治と選挙 . 東京大学出版会 .
- [3] B.S. Everitt, S. Landau & M. Leese (2001) Cluster Analysis 4th Edition, Arnold. London
- [4] F. Fehlen (1998), The cloud of candidates: exploring the political field, in *J. Visualization of Categorical Data*, ed. by Blasius and M. Greenacre, 159-170, Academic Press.
- [5] M. J. Greenacre (1988) Clustering the rows and columns of a contingency table. *Journal of Classification*, 5: 39-51.
- [6] L. Lebart (1994), Complementary use of correspondence analysis and cluster analysis. *Correspondence Analysis in the Social Sciences*, ed. by M . Greenacre and J . Blasius, 162-178, Academic Press.