

非対称データの解析法の総合的展望

齋藤堯幸

東京工業大学・社会理工研究科

非対称データ

心理学、社会心理学、社会学、マーケティング、生態学、社会科学、行動科学等のさまざまな学問分野で、2つの対象 j と k の2元関係の度合 o_{jk} を観測し、その関係が非対称である現象を扱う場合がしばしばある。 n 個の対象 $\{1, 2, \dots, j, \dots, n\}$ からなる一群を考える。2元関係 o_{jk} のデータを、行列表示して $O = (o_{jk})$ と記す。非対称とは、少なくとも1つのペア (j, k) について、 $o_{jk} \neq o_{kj}$ であることをいう。非対称性が、誤差に起因するものでなく、データに内在する重要な情報や構造とみなすときに、非対称性に焦点を当てた分析法は意味をもつ。構造とは、空間的構造、クラスター構造、ネットワーク構造など、現象解析の目的に応じて多様である。

たとえば心理学では、2つの刺激(対象) j, k のペアについて、その関連性を被験者に判断させる実験課題がある。 o_{jk} は類似性(similarity)、非類似性(dissimilarity)、混同率(confusion)、関連性(association)などとよばれる。その課題に応じて、データ行列を類似性行列、非類似性行列、混同行列、関連性行列とよぶ。

社会心理学の社会的交換(social exchange)の研究では、 o_{jk} は2つの資源 j を受け取り、資源 k を返した件数(交換の件数)を示す。社会学における、世代間の職業移動(occupation mobility)の場合、 o_{jk} は親世代の職業 j から子世代の職業 k への職業の移動件数を示す。人口の空間的移動(移住)の場合には、ある地点 j から地点 k への移動を示す。社会的ネットワークの場合、個人を単位とするネットワークでは、 o_{jk} は個人 j と k の間の社会的関係の度合(交流、友人、メンバーシップなど)を示す。組織を単位とするネットワークでは、 o_{jk} は組織 j と k の関係(取引、交渉など)を示す。

マーケティングにおける銘柄交換(brand switching)では o_{jk} はブランド j からブランド k への買い換えする消費者の人数、商品 j の代わりに商品 k を代替する人数などを示す。

生態学の例として、 o_{jk} は植生の状態 j から状態 k への遷移を表す。また植物 j と植物 k の空間的近接関係を示す。社会、経済の分野に目を向けると、情報、通信、金融取引、貿易などについても、主体(対象)間のフローが、通常は非対称関係にある。

手法とモデルの展望

非対称データは、多種多様な分野で実験、調査、観測または統計的記録により収集される。以下では、この類のデータを分析する手法やモデル展望する。

1. 一対比較法に関する非対称分析法

1) o_{jk} が対象間の優劣関係(0,1で記述)の場合、対象の順序構造(または順序尺度)や階層構造を探索する手法がある。

2) o_{jk} が対象間の選好比率や比較判断の確率を表す場合、統計的モデルに基づく分散分析法や心理的モデルに基づく尺度構成法がある。

2. 非対称データのグラフィカル表現技法

非対称行列 $O = (o_{jk})$ を、線形数学による行列分解の手法(特異値分解、スペクトル分解など)を利用して、データを空間的に表現する手法群がある。データの測度(類似性、非類似性、職業移動など)の性質には依存しない汎用的手法とみなせる。したがって、特

定の現象を説明するための手法ではなく、モデルを構築するためにデータを縮約する情報を提供する。

3．非対称多次元尺度構成法

多次元尺度構成法 (MDS) は元来、(非)類似性 o_{jk} の対象性を前提として、多次元空間に対象(刺激)の空間的表現を導出することを目的とした。非対称性を誤差ではなく、意味ある情報とみなす場合、非対称性をモデルに組み込み MDS を行う手法群がある。モデル構築にはデータの記述的な立場と、非対称現象を生起する過程を説明する立場、さらにその中間的な立場がある。 o_{jk} を表現するために内積を一般化したモデルや距離を一般化したモデルや手法がある。

4．非対称クラスター分析法

クラスター分析法は、多変量データを分析するものと2元関係のデータ $O = (o_{jk})$ を分析するものに大別される。後者に属する手法群は、元来 o_{jk} の対称性を前提として、対象を分類することを目的とした。非対称 MDS の発展と同じように、非対称性を有意義な情報として扱う場合に、非対称クラスター分析法がさまざまに発展した。階層的手法の場合、デンドログラム表現に非対称性を組み込むことが重要になる。従来のクラスター分析法の範囲での発展にとどまらず、2. や 3. の手法やモデルを組み合わせたものがある。また単相2元データに関する分析法に加えて、2相2元データの分析を転用したものがある。

5．非対称ネットワーク分析法

社会学、社会心理学などで扱う社会的現象は、ネットワークで記述されるものが多い。したがって、非対称ネットワークの分析は心理学よりも社会学で発展した。その場合、ネットワークに関する二元関係データ $O = (o_{jk})$ は、実験、観察や面接によって直接的に収集されたり、記録調査等によって間接的に収集される。ノードには個人、組織、国家などが対応する。心理的現象の場合、連想ネットワーク、意味ネットワーク、ニューラルネットワークなど、ノードには状態 (state) を対応させる。

社会的ネットワークをソシオマトリクスや関連性行列で記述すると、多くの場合に非対称行列になり、有向グラフ (directed graph) が対応する。2 値的 (binary) あるいは重みつき (valued) の有向グラフを対応させて、集団 (actor set) の中に社会的なクラスター (cohesive group) を探索する手法がある。対人関係を2値データで表現したソシオマトリクス $O = (o_{jk})$ の場合、有向グラフを構築し、対象間の関係を有向グラフ統計量を用いた確率モデルによって分析する方法もある。

心理的現象では連想記憶に関わる非対称データ $O = (o_{jk})$ から、刺激間のネットワークを構築し、そのノード間の距離を計量するスケーリング手法もある。

上記 1. から 5. の手法やモデルは、それぞれの立場から発展してきたが、それらを包括的に位置づけ総合的な解説書として、文献 [1] がある。

参考文献

- [1] Saito, T. & Yadohisa, H. (2004) Data Analysis of Asymmetric Structures: Advanced Approaches in Computational Statistics. Dekker, New York.