博士（工　学）　宿　久　　洋

学位論文題名

Studies on Asymmetric Multidimensional Scaling in Similarity Data Analysis

（類似性データ解析における非対称多次元尺度構成法に関する研究）

学位論文内容の要旨

多次元尺度法 (Multidimensional Scaling, 以下 MDS) 及びクラスタリング法は (非) 類似性データと呼ばれる対象間の関係の (非) 近接度を表すデータを解析するための手法で, 社会学, 行動科学, 工学をはじめ多くの分野で用いられている. 例えば, 国際貿易における輸出量, 車の買い換え, 電話の通話料, 研究雑誌間の引用の数等は典型的な (非) 類似性データであり, 枚挙に暇はない.

(非) 類似性データの非対称性とは, A から見た B の類似性が B から見た A のそれと一致しないことで, 上記の様なデータでは, その差異に重要な意味がある場合も少なくない. MDS は Torgerson (1952) 以来, 多くの研究者によって研究されてきたが, 従来の研究の多くがこの非対称性を誤差と考え無視するものであった. しかし, 近年この非対称に着目した研究が Gower (1977), Chino (1978), Weeks and Bentlar (1982), Sato and Kawaguchi (1990), Saito (1991), Zielman (1991), DeSarbo and Marrai (1992) 等により精力的に展開されている. しかしながら, 実際の応用分野での適用をみると非対称 MDS が従来の対称 MDS ほど積極的に利用されているとはいえないのが現状である. 著者はこの原因が手法の複雑さによる「解釈の困難」や非対称性に関する事前の知識を仮定することによる「制約」にあると考え, 本論文では非対称性に関する事前知識を仮定しない, 直感的な解釈が容易である新たな非対称 MDS の提案を行った. 一方, MDS の結果を具体的に応用する場合に, 結果のクラスタリングが有効である. そのためにクラスタリングの特性を解析するための方法として空間の伸縮に関して, 組合せ的クラスタリング法 (Lance and Williams, 1967) について考察した. この方法は, 1 つのクラスターに結合されたクラスター I, J と他のクラスター K との更新式 $d_{(IJ)K}$ が $d_{(IJ)K} = \alpha_i d_{IK} + \alpha_j d_{JK} + \beta d_{IJ} + \gamma |d_{IK} + d_{JK}| \quad (d_{IJ} < d_{IK} < d_{JK})$
で表される手法の総称として提案された. ただし, $\alpha_i, \alpha_j, \beta, \gamma$ は手法によって決定される適当な実数, d_{IJ}, d_{IK}, d_{JK} はそれぞれクラスター I と J, I と K, J と K の間の非類似度である. 組合せ的手法の中で, 結合過程においてクラスターが特定の方向に成長しやすく, 鎖状のクラスター (chain) ができる等の現象が起こることがある (Williams, Lambert, and Lance, 1966). これらの現象のことを Lance and Williams (1967) は空間のゆがみ (Space Distortion) と名付け, いくつかの実験によりこのゆがみを評価することを試みた. その後, DuBien and Warde (1979), Nakamura and Ohsumi (1990) によって更新式の満たす条件に基づき, 空間のゆがみを分類する提案がなされた. また, Chen and Van Ness (1994a, 1994b, 1996) によって空間のゆがみと Fisher and Van Ness (1971), Van Ness (1973) で導入されたクラスタリング法の許容性との関係が明らかにされた. これらの論文において, 空間のゆがみは用いる更新式のパラメータとの関係で論じられ, 各手法の用いる更新式が定義された

いずれの領域に属するかによってその手法を分類し, 空間を拡大, 縮小又は保存する手法であると解釈している. これらに対し本論文では, 空間のゆがみが用いる手法の違いのみではなく, 解析するデータによっても異なるのが自然であるという考えに基づき, 空間のゆがみを測る指標を提案し, 加えて, 新指標を用いたクラスタリング法の許容性及び新たな組み合わせ的クラスタリング法を提案した.

本論文は 6 章で構成され, 前半 2 章で非対称単相 2 元データを解析するための新たな非対称 MDS を提案し, 中半 2 章では前半で提案した手法を非対称 2 相 3 元データに適用するための拡張を行った. また, 後半 2 章では, 対称単相 2 元データを解析するための組み合わせ的クラスタリング法における問題点の 1 つである空間のゆがみについて議論し, そのゆがみを測る指標を提案し, 新指標を用いたクラスタリング法の許容性及び組み合わせ的クラスタリング法を提案した. 各章の詳細については以下のとおりである.

第 1 章では非対称性に関する基本的な考え方を述べるとともに, 非対称性を表現するための Vector model を提案し, 非対称ベクトルを定義した. 加えて, この非対称ベクトルを用いた対象間の様々な非対称構造を測る指標の提案を行った. 第 2 章では, 全体的な非対称構造を表現するための Vector field model を提案した. ここでは第 1 章の方法で得られた結果全体を 1 つのベクトル場と考え, その Scalar Potential を推定し視覚化することにより, 非対称構造をより見易く表示することを試みた. 第 3 章では第 1 章の Vector model を 2 相 3 元データに適用するために拡張した幾つかのモデルの提案を行った. 具体的には, Carroll and Chang (1970) の INDSCAL と Young (1984) の GEMSCAL に対応した 拡張 Vector model を提案した. 第 4 章では 非対称な 2 相 3 元データの解析法として, Ambrosi and Hansohm (1987) の Dynamic MDS を非対称に拡張した手法 (手法 1) と Sibson (1978) の Prosrustes analysis を用いる手法 (手法 2) の提案を行った. また, 解析結果の動的な表現法として Spline interpolation を用いた Time-space representation を提案した. 第 5 章では, 空間のゆがみを測る指標 Distortion Ratio を定義し, 人工例を用いて主な組み合わせ的手法の空間のゆがみに関する性質を論じた. 第 6 章では, 前章で定義した Distortion Ratio を用いてクラスタリング法の許容性を提案した. 加えて, 空間のゆがみを考慮した新たな組み合わせ的クラスタリング法した.

第 1 章で提案した非対称 MDS では非対称性が視覚的に対象と同じ空間に表現される. このことにより, 探索的に用いられることが多い多次元尺度法の利点, すなわち, 視覚的に結果を表現することができるという点を失うことなく, 非対称性を表現することが可能となった. また, ベクトルによる表現は, 第 2, 第 3, 第 4 章で論じたような拡張を容易にし, データの持つ非対称構造を様々な視点でとらえることができると考えられる. 次に, 第 5 章で提案した Distortion Ratio は空間のゆがみを数値で表す従来にはない指標であり, このことは, 第 6 章で行った応用, つまり, Distortion Ratio を利用した許容性やクラスタリング法の提案を可能にした. また, 第 6 章で提案した許容性は従来の空間のゆがみ, 単調性に関する許容性を特別な場合として含む一般的なものであり, クラスタリング法の性質を論じる上で有用な指標であると考えられる.

学位論文審査の要旨

主 査　教 授　佐 藤 義 治
副 査　教 授　新 保　勝
副 査　教 授　伊 達　惇
副 査　教 授　宮 腰 政 明
副 査　教 授　水 田 正 弘

学 位 論 文 題 名

Studies on Asymmetric Multidimensional Scaling in Similarity Data Analysis

(類似性データ解析における非対称多次元尺度構成法に関する研究)

類似性データとは，対象間の何らかの意味での類似している程度を測定したデータであり，工学, 社会学，心理学をはじめ様々な分野で扱われている．対象そのものに関わる量は直接観測できないが，対象間の関係に関わる量（類似性）は観測可能であるといった状況は数多く存在し, そのようなデータの解析手法の研究は重要な課題である．

本論文では，類似性データの解析手法である多次元尺度構成法 (MDS) について，データの非対称性に着目した理論的研究及び具体的な解析法の提案を行っている．また，MDS の結果を具体的に応用する場合に併用して用いられるクラスタリング法の特性についても空間の伸縮に着目した検討を行っている．

類似性データの非対称性とは，対象 A かみた対象 B の類似性が逆の場合のそれと異なるというものである．特に，ある地点間の情報や物の流量に関するものや行動科学等の分野において扱われるデータの多くが非対称なもので，そこに本質的な意味がある場合も少なくない．近年，この非対称性に着目した MDS の研究が精力的に展開されているが，実際の応用分野では対称な類似性に対する MDS ほど利用されているわけではない．著者はこの原因が手法の複雑さによる「解釈の困難さ」や非対称性に関する事前の知識を仮定することによる「制約」にあると考え，本論文では非対称性に関する事前知識を仮定しない，直感的な解釈が容易である新たな非対称 MDS の提案を行っている．

本論文は序論とそれに続く 6 章から構成され, 各章における成果はつぎの通りである．

第 1 章では，非対称性に関する基本的な考え方を述べるとともに，非対称性を表現するための Vector model を考案し，非対称性ベクトルを定義している．さらに，解析結果全体を 1 つの点系と考え，系全体の持つ並行的，回転的，膨張的非対称性に関する各指標を導出している．このような指標の一部は既存の手法で個別に扱われてきたが，本手法により統一的な見方が可能になったと考えられる．また，系がいくつかのクラスターに分かれた場合についても，個々のクラスターに対して各指標を定義し，その性質について議論している．ここで提案されている手法では，対象自体を表す点と対象間の非対称性を表すベクトルが同一の

空間に表現され，従来の手法と比較すると，より直感的な解釈が可能であると考えられる．また，事前に非対称性に関する知識を仮定しない点は特筆すべき点である．

第2章では，第1章で提案した手法で得られた結果全体を1つのベクトル場と考え，そのScalar Potentialを推定し，視覚化することにより，結果の全体像をより見易く表示する方法を示している．ここでは，非対称ベクトルがScalar Potentialの勾配とVector Potentialの回転の和で書けると考え，前者をその点を含む単位格子上の立体四辺形の「勾配」で近似し，後者をランダムベクトルのようにみなしている．このような構造を仮定して，各点それぞれにおける勾配の近似の良さと全体にわたる平滑条件の満足度を同時に高めるようなScalar Potentialの推定法を確立している．Scalar Potentialを視覚化することにより，解析結果のより大域的な解釈が可能となり，全体的な特徴抽出に有用である．

第3章では，複数個の非対称類似性行列がデータとして与えられる場合について，複数の対称類似性行列が与えられる場合に用いられる2つの手法，Chang (1970) のINDSCALおよびYoung (1984) のGEMSCALを著者の提案しているVector modelの考え方に基づき拡張している．このことは，ほとんどの類似性データの解析において本論文で提案している考え方に基づく非対称性の表現が可能であることを意味し，実用上, 有効な方法である．

第4章では，複数のデータ行列の中でも特に，経時的に観測されたデータが与えられた場合について，Ambrosi and Hansohm (1987) のDynamic MDSを非対称に拡張した手法 とSibson (1978) のProsrustes analysisを用いる手法 の提案を行っている．従来，このようなデータに対する非対称ＭＤＳの適用の試みはほとんどなく，ここでの取り組みは大変注目すべきものである．また，上記の手法による解析結果の動的な表現法としてSpline補完法を用いたTime-space表現を考案している．

第5章では，組み合わせ的クラスタリング法の解析結果にみられる「空間のゆがみ」という現象に着目し，数理的な評価を行っている．空間のゆがみは早くから知られていた現象ではあったが，数理的な定義付けは近年になってから与えられた．著者は，その概念を一般的に拡張したゆがみ度と呼ばれる指標を提案している．ゆがみ度は利用者が解析手法を選択する場合や解析結果を判断する場合の新たな指標として有用であると考えられ，ゆがみ度と他の手法選択のための指標を併用することで，より客観的な判断が可能になると考えられる．

第6章では，ゆがみ度を用いてクラスタリング法の許容性と新たなクラスタリング法の提案を行っている．クラスタリング法における許容性は，手法の選択において重要な概念の一つである．著者は空間のゆがみに関するいくつかの許容性を提案し，これらの許容性と主な組合せ的クラスタリング法との関係を明らかにしている．これを用いることにより，解析者の用いているクラスタリング法が空間のゆがみの意味でどのような性質を持つかが明確に示され，解析結果を解釈する上で重要な情報を得ることができる．

これを要するに，著者は，非対称性を有する類似性データに対して, Vector Modelという新しい概念を用いた多次元尺度構成法を開発し，その有効性を示している．またそれに関連して, クラスター分析におけるいくつかの数理的な新知見を得たものであり，情報解析学および計算機統計学の進歩に寄与するところ大なるものがある．

よって，著者は北海道大学博士（工学）の学位を授与される資格あるものと認める．