

**科学研究費助成事業 研究成果報告書**

平成 28 年 6 月 19 日現在

機関番号：22604

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25730019

研究課題名(和文)多相多元データと非対称データの特性に即した解析法選択のための方法と新解析法の提案

研究課題名(英文) Study of selection of method and suggestion of new method considering attribute of multi-mode multi-way data and asymmetry data

研究代表者

中山 厚穂 (Nakayama, Atsuhō)

首都大学東京・社会(科)学研究科・准教授

研究者番号：60434198

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：様々な分野で、複数の情報が組み合わされた構造や形式が複雑なデータを分析することへの需要が高まっており、データの性質に適した解析を行うことのできる多変量解析手法が求められている。そのため、データの構造や形式に適した解析を行うための手法についての研究が進められているが、これらの手法が一般の応用分野に普及しているとはいいがたく、どのような性質のデータにはどのような解析法を適用すべきなのかという指針は確立されていない。以上から、本研究課題では多相多元データと非対称データの構造や形式に即した解析法の選択法の構築と新解析法の提案を目的として研究を行う。

研究成果の概要(英文)：testNeeds for the analysis of data with complex structure and form that consist of a lot of information in many fields have increased. The models that can appropriately analyze the attribute of these data would have been requested. The models to appropriately analyze the data considering the structure and the form have been proposed. However, these models have yet to become widely used and there are no criteria to select the analysis that is appropriate for the structure and form of data. In this study, we proposed the framework of the selection of the appropriate model and the new model to analyze multi-way multi-mode data asymmetric data considering the attribute of the structure and form of the data.

研究分野：統計科学

キーワード：単相多元データ 非対称データ 多次元尺度構成法 対数線形モデル

### 1. 研究開始当初の背景

社会現象や心理現象を扱う様々な学問分野あるいは応用分野で、調査項目、変数、回答者、データ収集時点など複数の情報が組み合わせられた構造や形式が複雑なデータが入手可能となっている。データの構造や形式は、データの相と元の数によって分類することができる(Carroll & Arabie, 1980)。相は、一組の対象(調査項目、変数、回答者、データ収集時点など)を意味し、相の数は、データに相がいくつ(何種類)存在しているのかを示す。元の数は、相がデータで合計何個組み合わせられているのかを表す。

多相多元データの分析にあたっては各データの構造や形式に適した手法で分析することが望ましい。しかし、回答者×調査項目×調査時点からなる3相3元データを分析する際に、1つの元の圧縮やデータを分割して2相2元データとした上で分析するというようにデータを分析方法に適した形式に変換した上で分析されることが多い。そして、このような圧縮や分割の是非を吟味してから分析を行う場合もあるが、多くの場合にはそのような吟味は行われぬ。多相多元データの元(もしくは相)を圧縮することは、その元(相)がもつ情報を利用しないこと、あるいは、それ以外の元(相)が表すデータが圧縮された元(相)について同一、つまり、圧縮された元(相)における変動は誤差であるとみなすことになる。また、2元データなどに分割し、各々のデータを別個に分析すると、各データの分析結果が独立に得られ、個々の分析結果には関連がなく、結果の解釈が難しいことが多い。したがって、多相多元データの分析には、そのデータの構造や形式に適した手法を用いて、多相多元データを圧縮したり分割したりせずに、そのまま多相多元データとして分析することが望ましい。これは他のデータの構造や形式(例えば、非対称データをそのまま分析せずに対称化して分析するような場合)においても共通した問題である。一方で、データの構造や形式に適した手法によりすべての情報を反映した解析を行うことができたとしても、その結果をすべて正しく理解できるという保証はない。社会現象や心理現象をより単純なモデルで表現をした方が分析結果を一般化しやすい。そこで、データの構造や形式に即した解析法のための方法論の構築が重要ではないかと考える。また、近年では、より構造や形式が複雑なデータに適用するための手法の開発提案がなされている。しかし、一般の応用分野に普及しているとはいえず、さらに、取り扱うデータに様々な仮定がおかれている場合もあり、データ解析のニーズに応えているとはいえない。

### 2. 研究の目的

本研究ではデータの構造や形式に即した適切な解析法の選択のための方法論と新解

析法の提案を目的として研究を行った。データの構造や形式については多相多元データ、非対称データに注目して研究を進めた。

多相多元データや非対称データの分析には、そのデータの構造や形式に適した手法を用いて、多相多元データを単相2元データや2相2元データなどに圧縮したり分割したりせずに、またデータのもつ非対称性の情報を考慮した上で分析するのが望ましい。しかし、多相多元データや非対称データを分析するのに適した手法は、これまでに一般的に用いられてきた単相2元データや2相2元データ、対称データを分析するための手法と比べるとより複雑なモデルとなる。モデルがより複雑となれば、分析結果の解釈が困難となることや分析結果が一定のデータの特性に即したものとなり一般化が難しくなるなどの問題も想定できる。よって、本研究では多相多元データと非対称データというデータの構造や形式における適切な解析法の選択のための方法論を構築し、単純なモデルとより複雑なモデルのいずれかで分析すべきなのかというモデル選択についての指針を示すことを目指す。

多相多元データにおいて、データを圧縮・分割することで対象間の関係が損なわれるかどうかについて検討するために、対数線形モデルにより各相や元における交互作用の関係性を明らかとし、圧縮や分割を行うことの妥当性について検証する。

非対称データにおいては2元分割表の対称性検定(対称性の適合度検定、周辺対称性の検定、準対称性の検定)を多元データに対張した検定方法を提案することで、非対称モデルで分析すべきなのかどうか、その妥当性について検証する。

そして、データの構造や形式における適切な解析法の選択のための方法論についての研究を踏まえ、多相多元データと非対称データの構造や形式に即した新たな解析手法の提案を目指す。

### 3. 研究の方法

多相多元データの構造や形式に即した解析法の選択のための方法論について研究では、Nakayama (2013)やNakayama (2014)の研究に代表されるような、各元の交互作用を考慮することで、データを圧縮・分割することの妥当性を吟味する研究と中山・増田・鶴見(2015)のようなデータの構造や形式に即した解析法についての研究を行った。Nakayama (2013)では、より元の数が多い高次のモデルと元の数が少ない低次のモデルによる分析結果を一致係数や順位相関係数などを用いてより比較検討することで、元を圧縮・分割することが妥当性を検証した。また、Nakayama (2014)では対数線形モデルにより、各相の交互作用を考慮した上で、元を圧縮・

分割することが妥当かを検証するための方法論を提案した。中山・増田・鶴見(2015)ではTwitterの新製品についてのTwitter上の書き込みから作成した期間×書き込み×単語からなるスパースな3相3元データの有するデータの構造や形式に即した解析法についての研究を行った。

非対称データの構造や形式に即した解析法の選択のための方法論について研究では、Nakayama, Tsurumi, and Okada (2013)に代表されるような2元分割表の対称性検定(対称性の適合度検定, 周辺対称性の検定)を単相3元データに対張した検定方法を提案し、非対称モデルで分析すべきなのかかどうかという、非対称データの構造や形式に即した解析法の選択のための提言を行った。

また、本研究課題において提案を行った方法論を実データに適用し応用分野への適応可能性について検討した。

#### 4. 研究成果

多相多元データの構造や形式に即した解析法の選択のための方法論について研究においては、Nakayama (2013)や Nakayama (2014)などの研究により、高次のモデルとより低次のモデルによる分析結果を比較検討し、また、各元の交互作用を考慮することで、データを圧縮・分割することが妥当かを吟味した上で分析を行うことの必要性と重要性を示すことができた。中山・増田・鶴見(2015)ではデータの構造や形式に即した解析法についての研究を行い、データの特性について吟味した上で分析を行うことの重要性を示すことができた。

非対称データの構造や形式に即した解析法の選択のための方法論について研究を行った Nakayama, Tsurumi, and Okada (2013)では単相3元非対称データに対して、対称性の適合度検定, 周辺対称性の検定を実施することで、非対称モデルで分析すべきなのかどうかというモデル選択のための指針を示し、その研究成果から提案した検定法によるモデル選択の有効性を示した。

#### 引用文献

- Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 31, pp. 607-649). Palo Alto, CA: Annual Reviews.
- Nakayama, A. (2014). Evaluating the Necessity of a Triadic Distance Model [summary]. Abstracts of Second European Conference on Data Analysis (ECDA),

Jacobs University Bremen, Germany (3 July 2014).

Nakayama, A. (2013). Analysis of Conditional and Marginal Association in One-Mode Three-way Proximity Data. In W. Gaul, A. Geyer-Schultz, Y. Baba and A. Okada (Eds.), *German-Japanese Interchange of Data Analysis Results* (pp. 135-146). Berlin, Springer.

Nakayama, A. Tsurumi, H., and Okada, A. (2013). A Symmetry Test for One-Mode Three-Way Proximity Data. In W. Gaul, A. Geyer-Schultz, Y. Baba and A. Okada (Eds.), *German-Japanese Interchange of Data Analysis Results* (pp. 125-134). Berlin, Springer.

中山厚穂・増田純也・鶴見裕之 (2015). 製品時系列を考慮した Twitter 上のトピック分類 データ分析の理論と応用(査読有), Vol.4, 17-41.

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 9 件)

Nakayama, A (2016). Evaluating the Necessity of a Triadic Distance Model. In A. F. X. Wilhelm, and H. A. Kestler (Eds.), *Analysis of Large and Complex Data* (査読有), Springer. DOI: 10.1007/978-3-319-25226-1

中山厚穂・増田純也・鶴見裕之 (2015). 製品時系列を考慮した Twitter 上のトピック分類 データ分析の理論と応用(査読有), Vol.4, 17-41.

Iguchi, C., Hayashi T., and Nakayama A. (2014). Global Innovation and R&D for Knowledge Creation: The Case of P&G, Unilever and Kao. In P. Strom, H. Dolles, C. and R. Middel (Eds.), *Asian Inward and Outward FDI: New Challenges in the Global Economy* (pp.65-86) (査読有), Basingstoke, Palgrave. DOI: 10.1057/9781137312211

Iguchi, C., Hayashi T., and Nakayama A. (2014). The Innovation System of Culture-Specific MNEs: The Effects of Diversified and Geographically Dispersed Knowledge Sourcing Mechanism, *Keio Business Review* (査読有), No. 48, 25-39.

Nakayama, A. (2013). Analysis of Conditional and Marginal Association in One-Mode Three-way Proximity Data.

In W. Gaul, A. Geyer-Schultz, Y. Baba and A. Okada (Eds.), German-Japanese Interchange of Data Analysis Results (pp. 135-146) (査読有). Berlin, Springer.  
DOI: 10.1007/978-3-319-01264-3\_12

Nakayama, A. Tsurumi, H., and Okada, A. (2013). A Symmetry Test for One-Mode Three-Way Proximity Data. In W. Gaul, A. Geyer-Schultz, Y. Baba and A. Okada (Eds.), German-Japanese Interchange of Data Analysis Results (pp. 125-134) (査読有). Berlin, Springer.  
DOI: 10.1007/978-3-319-01264-3\_12

鶴見裕之・増田純也・中山厚穂 (2013). マーケティングにおける SNS 上のテキスト・データ活用の可能性と限界. マーケティングジャーナル(査読無), Vol.138, 38-54.

鶴見裕之・増田純也・中山厚穂 (2013). 商品に関する Twitter 上のコミュニケーションと販売実績の関連性分析. オペレーションズ・リサーチ(査読無), Vol.58, No8, 436-441.

鶴見裕之・小木しのぶ・中山厚穂 (2013). 誌上セミナー「テキストマイニング」(3) 流通業・製造業のためのテキスト・データ活用, 流通情報(査読無), No.502 (Vol.44, No.7), 82-95.

[学会発表](計 16 件)

Nakayama, A. (2015). Classification and Prediction of Topics on Social Media considering Temporal Variation [summary]. Abstracts of the 9th Conference of the Asian Regional Section of the IASC, Singapore (Singapore), December 17, 2015, Invited session.

中山厚穂 (2015). 対象集団類型化のための統計モデル～オーバーラッピングクラスタリングの基礎と活用事例. 5 回スポーツデータアナリティクス基礎講座, 201, 2015 年 7 月 19 日, 慶應義塾大学 (神奈川県・横浜市).

Nakayama, A. (2015). The Classification and Visualization of Twitter Trending Topics Considering Time Series Variation [summary]. Abstracts of the 2015 Conference of the International Federation of Classification Societies (IFCS), Italy (Bologna), July 7, 2015, Invited

session.

鶴見裕之・増田純也・中山厚穂 (2015). Web コミュニケーション・データを活用した消費者理解, 日本マーケティング・サイエンス学会第 97 回研究大会, 2015 年 6 月 14 日, 大阪府立大学 (大阪府・堺市).

中山厚穂 (2015). クラスタリング・セグメンテーションとその応用, 日本行動計量学会第 17 回春の合宿セミナー, 2015 年 3 月 8 日, 東京大学 (東京都・目黒区).

中山厚穂 (2015). オーバーラッピング・クラスタリング入門, 日本分類学会データ分析セミナー, 2015 年 3 月 2 日, 帝京大学 (東京都・千代田区).

鶴見裕之・増田純也・中山厚穂 (2014). Twitter 上の"消費者の声"の内容と販売実績の関係性分析, 日本マーケティング・サイエンス学会 96 回大会, 2014 年 11 月 30 日, 筑波大学 (東京都・文京区).

出口慎二・中山厚穂・高崎祐哉 (2014). 大規模データのクラスタリング, 日本行動計量学会第 42 回大会, 2014 年 9 月 3 日, 東北大学 (宮城県・仙台市).

鶴見裕之・増田純也・中山厚穂 (2014). Twitter 上のコミュニケーション内容と販売実績の関連性を考慮した顧客開発施策, 日本行動計量学会第 42 回大会, 2014 年 9 月 3 日, 東北大学 (宮城県・仙台市).

中山厚穂・下川敏雄・酒折文武・竹村彰通 (2014). JMOOC「統計学講座」の開発, 日本行動計量学会第 42 回大会, 2014 年 9 月 3 日, 東北大学 (宮城県・仙台市).

Nakayama, A. (2014). Evaluating the Necessity of a Triadic Distance Model [summary]. Abstracts of Second European Conference on Data Analysis (ECDA), Germany (Bremen), 3 July 2014.

鶴見裕之・増田純也・中山厚穂 (2014). Twitter 上のコミュニケーション内容を考慮した販売実績データ分析, 日本マーケティング・サイエンス学会 95 回大会, 2014 年 6 月 22 日, 関西学院大学 (兵庫県・西宮市).

中山厚穂・鶴見裕之・増田純也 (2013). 製品時系列にともなう出現単語変化に注目した Twitter 上のトピック分類, 日本行

動計量学会第 41 回大会, 2013 年 9 月 4 日, 東邦大学 (千葉県・船橋市).

鶴見裕之・増田純也・中山厚穂 (2013). Web コミュニケーションとマスコミュニケーションを加味した販売実績データの分析, 日本行動計量学会第 41 回大会, 2013 年 9 月 4 日, 東邦大学 (千葉県・船橋市).

Nakayama, A., Tsurumi, H., and Masuda, J. (2013). Classification of Topics on Twitter in Consideration of Time Series Variation [summary]. Abstracts of the 2013 conference of the International Federation of Classification Societies (IFCS), Netherland (Tilburg), 17 July 2013.

Hayashi, T., Nakayama, A. and Iguchi, C (2013). Strategic Knowledge Creation and the Management of Diversities - Comparative Analysis of Kao Corporation and P&G [summary]. Abstracts of the Association of Japanese Business Studies (AJBS), Turkey (Istanbul), 3 July 2013.

鶴見裕之・増田純也・中山厚穂 (2013). Twitter 上の書き込みデータを活用した販売実績データ分析, 日本マーケティング・サイエンス学会 93 回大会, 2013 年 6 月 9 日, 長崎大学 (長崎県・長崎市).

〔図書〕(計 1 件)

竹村彰通・下川敏雄・酒折文武・中山厚穂 (2014). 統計学 データ分析の基礎 オフィシャルスタディノート, 日本統計協会.

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

中山 厚穂 (Nakayama Atsuho)

首都大学東京社会科学部研究科・准教授

研究者番号: 60434198