

平成 30 年 6 月 12 日現在

機関番号：34315

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26380658

研究課題名(和文) 社会学的データ分析へのランダム行列理論の応用にかんする研究

研究課題名(英文) Application of Random Matrix Theory to sociological data analysis

研究代表者

中井 美樹 (NAKAI, Miki)

立命館大学・産業社会学部・教授

研究者番号：00241282

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本研究は、ランダム行列理論を社会学的データ分析に応用し、幾何学的フレームワークの応用により精緻な分析モデルに基づいた社会調査データ分析の新たな手法を提案することを目指すものである。研究期間には、第一に、社会調査データを扱う際の課題の1つであるカテゴリカル・データと連続変数を分析できる幾何学的フレームを提案した。第二に、ランダム行列理論に基づく手法を大規模社会調査データの主成分分析に応用した。第三に、欠損データ(missing data)を含むデータセット分析に本手法を拡張させることを試みた。

研究成果の概要(英文)：During the 2014-2017 period, our research focused on applying mathematical techniques from Random Matrix Theory (RMT) to central problems in sociological data analysis. First, we developed a novel geometrical framework for correlation matrices of heterogeneous (i.e. categorical and continuous) sets of variables. Second, we applied RMT to the Principal Component Analysis of the SSM2005 dataset, within the geometrical framework. Third, we extended our analysis to the Missing Data problem, and developed a new method that maximizes the number of non-missing data entries after list-wise deletion. Finally, we produced algorithms in Matlab/Octave and R, and disseminated results at international conferences.

研究分野：社会学

キーワード：ランダム行列理論 共分散行列 カテゴリカル・データ 社会学データ 欠損値データ

1. 研究開始当初の背景

社会学研究においては、収集された社会調査データの分析に基づく経験的・実証的研究によって種々の社会現象の理解やメカニズムを解明し、社会学理論を構築・展開してきた。その際に活用される社会調査データに特有の、分析に伴う課題がいくつかある。例えば、社会学的な事象の尺度や測定の際にはカテゴリカル変数(名義尺度、順序尺度など)によってデータが収集されることが多く、カテゴリカル・データと連続データを同時に分析する必要があること、さまざまな理由により回答の得られないデータ(欠損・欠測データ、不完全データ)が生じること、複雑な事象の関連性を解明するため多くの変数を同時に分析する必要がある一方で、近年の調査環境の悪化による回答率の低下に伴いサンプルサイズが十分に大きい社会調査を実施する困難さが増していること、などである。社会調査データを扱ううえでのこうした諸課題に対処する手法やモデルの開発・応用が社会学研究や行動計量学などの領域において進められ、あわせて調査設計に関する研究も進められてきた。その際にはしばしば、広範な社会諸科学領域や自然科学の諸領域で有用とされる枠組みやアイデア・手法も社会学的データの分析に様々な利用可能性を持つと考えられ、応用されてきた。その一方で、近年、社会科学分野(金融工学、情報学、ネットワーク分析など)で応用研究が広がりつつあるランダム行列理論は、これまで社会学領域での社会調査データ、いわゆる標本調査データの分析には適用されてこなかった。

ランダム行列理論はランダム行列、すなわち確率変数を要素に持つ行列を扱い種々の現象のモデル化を行う。従来の研究から、ランダム行列理論の適用によってデータの「誤差」と「情報(=真の値)」の部分の識別が可能となることが示されてきた。また近年はカテゴリカル・データを分析対象とする場合にも相関行列に基づく手法を応用した分析が提案されている。したがって、この枠組みを社会学データに応用することによって種々の課題への対処が可能となり、結果として既存の手法に基づく知見よりもより確かで新たな社会学的知見が導ける可能性があると考えた。

上記の経緯より、ランダム行列理論を適用する新たな手法を社会学における社会調査データ分析に応用して、それによって社会学的データに伴う課題を克服し、より精緻な分析モデルに基づいた社会調査データ分析の新たな手法を提案し、よりの確で深い社会学的インプリケーションを得ることを目指して、まず学際的な研究者による研究組織を構築することで研究目的を実現するため、科学研究費補助金(基盤研究(C)「社会調査の分析モデルにおけるランダム行列理論の応用」、平成23~平成25年度)の採択・交付を受けて研究を開始した。そこで構築された研究組

織を基盤として研究を継続させ、得られたアイデアを発展させ社会学的方法論を深化させる必要があると考えるにいたった。

2. 研究の目的

本研究の目的は、社会調査データ分析へのランダム行列の応用により、具体的には主として以下の2つの課題に取り組むことである。第一に、ランダム行列理論(Random Matrix Theory)を、社会学における大規模社会調査データ分析に応用すること、具体的には、多くの変数を含む社会学的データセット分析に新たな幾何学的手法を応用することでより深い社会学的インプリケーションを導くことである。社会学的な大規模な調査データセットは多数の変数を含む一方で、変数の数に対してオブザベーション数が必ずしも十分に多くないといった分析の問題点があると言われ、このように行列の次元が大きい場合の推定に伴う問題の一つに、統計的誤差(標本誤差)が大きくなることがある。本研究では、これまで社会学的データセットの分析への応用が試みられてこなかったランダム行列理論を用いることで、多変量のデータに含まれる誤差を除去し、それによって推定の精度をあげるといった手法の可能性について検討する。

第二に、欠損値を含む社会学的データセットの頑健な分析手法を提案することである。社会調査のデータセットには、しばしば、様々な理由により「無回答」いわゆる不完全データへの対処という問題が生じる。実証的研究の現実の文脈では種々の欠損値データに対し緻密に対処すべきであるとする議論がRubinやAllisonなどによって従来からなされている(Little and Rubin 2002, Allison 2000)。本研究ではランダム行列理論のアプローチからこの問題に対処する手法を考案する。

3. 研究の方法

本研究では、H23~25年度基盤研究(C)での、行列理論と線形代数を社会調査データの分析モデルに応用した新たな手法の提案などの成果に基づき、そこで編成した研究組織を維持しながら研究を発展・深化させるべく研究を進めることとした。

2.で述べた研究目的を達成するための研究体制として、主として2つの研究部門を編成した。これらは第一に、「A.社会学研究法における幾何学的枠組みの応用およびランダム行列理論の応用」、「B.欠損値を含むデータに関する欠損メカニズムの理論的研究や既存研究の整理および手法の応用」からなる。また研究成果を発表し議論を深めるため、研究会を年に数回開催して集中的な議論を行いながら研究計画の効率的な遂行を行った。全体を通じて研究代表者が統括をつとめ、研究協力者との有益な連携を通じて全体テーマを学際的・総合的に推進した。

A. については、社会科学での応用が比較的新しいランダム行列理論について、理論的研究や既存の先駆的応用研究の概要について文献の整理を行い、報告と検討を行った。またそれを可能にする統計ソフトウェアについて知識・情報を収集し最近の研究動向について検討を行った。これと並行して、とりわけ社会学領域での社会調査データで多くを占めるカテゴリカル・データ(名義、順序尺度)を含むデータセットを扱うための分析手法にランダム行列理論を拡張させることを試みた。第二の研究部門Bは主として研究協力者を中心に推進し、加えて分析モデルのプログラミングも責任を持って研究を進める。海外の研究協力者とは、電子メールなどオンラインによる情報の共有を図るとともに、年に数回程度の研究会を実施し、互いに知見の共有とフィードバックを行いながら研究計画の効果的な遂行をめざした。

本研究ではすでに収集された社会調査データや、二次データに対してランダム行列モデルを応用し、ランダム行列理論の社会学的研究における実証的有効性を提示することとしていた。活用したデータセットは、例えば、日本での代表的な社会調査である「社会階層と社会移動全国調査(SSM調査)」データや「階層と社会意識全国調査(SSP調査)」データなどである。

4. 研究成果

(1) 研究会の開催

本研究は、研究代表者が責任を持って研究を遂行することとあわせて、社会学分野における研究での社会調査データ分析にこれまで応用されることがなかったランダム行列理論について、経済物理学や生態学等での応用研究をレビュー・検討し応用するため、社会科学領域でのランダム行列モデルの応用研究に精通した海外の研究者の協力のもとで、研究会での研究報告や検討などを通じて得られた新たな多様な研究成果を蓄積しながら学際的研究を推進してきた。研究会を開催し最新の研究情報を得るとともに研究交流と今後の継続的な研究体制の構築を図った。研究会の開催を通じて以下のような課題を整理・検討した。

[平成 29 年度]

第 1 回研究会 (H29. 9) CLADAG2017 年大会の発表打合せ

[平成 28 年度]

第 1 回研究会 (H28. 2) 欠損データの問題と幾何学的分析モデルの応用の検討(1)

第 2 回研究会 (H28. 5) 欠損データの問題と幾何学的分析モデルの応用の検討(2)

第 3 回研究会 (H29. 2) 研究成果のとりまとめと公表に向けた準備

[平成 27 年度]

第 1 回研究会 (H27. 7) 国際分類学会連合大会 (IFCS) 2015 年大会の発表打合せ

第 2 回研究会 (H27. 9) 研究成果のとりまとめに向けた準備

第 3 回研究会 (H27. 11) 社会学的データにおける欠損データの課題の整理

[平成 26 年度]

第 1 回研究会 (H26.5) 社会学的データの解析手法における課題の整理

2015 年、2017 年の研究発表とそれに基づく議論を受け、さらに継続的に本研究を発展・深化させる必要性が研究代表者および研究協力者の間で認識されるに至った。

(2) 国際学会大会への参加と成果発表

本研究期間を通じて達成した研究成果は多く、主要な成果は学会報告の形や論文などにより発表した。以下は、本研究期間を通じて達成した研究成果のうち主なものである。

研究成果の中間報告として、2015 年 7 月 6 ~ 8 日にポロニヤ大学で開催された Conference of the International Federation of Classification Societies (IFCS2015) において研究発表を行った。本研究での提案手法であるランダム行列理論にもとづくデータ分析手法を、大規模社会調査データセットに応用して、いわゆる“ノイズ”を除去した相関行列を再構築し、これに基づく種々の分析(主成分分析など)を進め、従来の手法に基づく結果との類似点・相違点や手法の課題を明らかにしようとした。

また非対称データ分析を用いた成果報告においては、非対称クラスター分析を用いた分析結果と多次元尺度構成法による結果とを比較し、併用が推奨されるこれらの手法から共通に得られる知見と個別に得られる知見があることを例証した。手法の応用領域および欠損データの対処手法についてなど多様なフィードバックを得ることができ、本研究の意義とその限界をあわせて示すとともに今後の研究を深化させることに繋がった。また欠損値への対処手法についても、学会報告において多くの研究者と議論を深め、既存アプリケーションソフトおよび既に提案されたパッケージの理解を深める機会を持った。

2017 年 8 月 8 ~ 10 日に、東海大学高輪キャンパスで開催された Conference of the International Federation of Classification Societies (IFCS2017) において、複数年次にわたる繰り返しのクロスセクション調査データを扱いながら、データの欠損のために生じる可能性のあるバイアスについて考慮した応用的分析を行った。調査年次によって調査データの回収率や欠損データの発生状況が変化し、欠測サンプルによる推定のバイアスが生じる可能性があるため、サンプリングウェイトを考慮した分析に

について検討した。この研究はミラノ・ピッコカ大学の研究者との新たな共同研究を開始するアイデアに結びき、新たな研究基盤を構築し得たことは大きな成果である。

2017年9月13～15日にミラノ・ピッコカ大学で開催された The 11th Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society (CLADAG2017)において、幾何学的手法を欠損値処理への対処にも応用・発展させ、「階層と社会意識全国調査 (SSP2015)」データの分析に応用した研究成果の発表を行った。研究開始当初には本課題の研究期間はH26～28年度を予定していたが、研究成果の取りまとめと報告の準備に取り組む中で「欠測値を含む調査データ」の追加的な分析の必要が生じた。このため研究期間を1年間延長してデータ解析に取り組み、成果の発表をH29年度に行うこととした。

社会調査データにほぼ必然的に発生しうる欠損値の問題にかんし、欠損値が存在するデータセットを扱う際に従来しばしば用いられてきたリストワイズ削除法を適用する場合の最適化の手法について開発・提案を行った。新規のアルゴリズムにより、回答者と変数について最適なデータセットを選択して分析対象とすることができる。さらにこれを実行するためのアルゴリズムを Matlab および R により作成した。この成果、すなわち拡張的な新たな手法の定式化と社会調査データ分析への応用については、英文論文としてまとめ海外の社会学ジャーナルに投稿している。

また本研究を進める過程で、社会調査データに伴う欠損値データに対処する手法について成果報告を契機として国際学会において様々な専門領域の研究者と慎重に議論を行う中で新たな課題に気づいた点も成果といえる。例えば、欠損値などの不完全データの問題に対処する際には、変数選択を主観的基準や客観的基準に基づいて進めることが一般的だが、こうした判断をいかにモデル化するか、などさらに継続的に本研究を発展・深化させる必要性が研究代表者および研究協力者の間で認識されるに至った。

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 2 件)

Nakai, Miki, "Social Differentiation of Cultural Taste and Practice in Contemporary Japan: Nonhierarchical Asymmetric Cluster Analysis," Palumbo, Francesco, Montanari, Angela, Vichi, Maurizio (Eds.) *Data Science: Innovative Developments in Data Analysis and Clustering*, Springer, Heidelberg-Berlin, 2017, 149-159, 査読有

DOI: 10.1007/978-3-319-55723-6

Vernizzi, Graziano. and Miki Nakai, "A Geometrical Framework for Covariance Matrices of Continuous and Categorical Variables," *Sociological Methods and Research*, 査読有, 44(1), 2015, 48-79

DOI: 10.1177/0049124114543243.

〔学会発表〕(計 5 件)

Miki Nakai, "Changes in Couples' Breadwinning Patterns and Wife's Economic Role in Japan," 11th Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society (CLADAG), 2017.

Vernizzi, Graziano and Miki Nakai, "Optimization of the Listwise Deletion Method," 11th Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society (CLADAG), 2017.

Nakai, Miki, "Changes in the Gendered Division of Labor and Women's Economic Contributions within Japanese Couples," Conference of the International Federation of Classification Societies, 2017.

Vernizzi, Graziano. and Miki Nakai, "Random Matrix Theory analysis of the 2005 Japanese national survey on social stratification and mobility," Conference of the International Federation of Classification Societies, 2015.

Nakai, Miki, "Social Differentiation of Cultural Taste and Practice in Contemporary Japan: Nonhierarchical Asymmetric Cluster Analysis," Conference of the International Federation of Classification Societies, 2015.

〔図書〕(計 2 件)

中井美樹、「階級的消費」日本社会学会理論応用事典刊行委員会『社会学理論応用事典』丸善出版、2016年、440-441頁

Nakai, Miki, "Sociology of Consumption" in James D. Wright(Ed.) *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition. Elsevier, Oxford, 2015, 757-761

6. 研究組織

(1)研究代表者

中井 美樹 (NAKAI, Miki)

立命館大学・産業社会学部・教授

研究者番号：00241282

(2)研究分担者
なし

(3)連携研究者
なし

(4)研究協力者
Graziano Vernizzi
Siena College • Department of Physics and
Astronomy • Professor