

令和 4 年 6 月 13 日現在

機関番号：31308

研究種目：研究活動スタート支援

研究期間：2020～2021

課題番号：20K22125

研究課題名（和文）機械学習の識別可能性の探求と消費者行動分析への応用

研究課題名（英文）Investigating the identifiability of machine learning and its application to consumer behavior analysis

研究代表者

佐藤 平国（Sato, Toshikuni）

石巻専修大学・経営学部・助教

研究者番号：10878804

交付決定額（研究期間全体）：（直接経費） 2,000,000円

研究成果の概要（和文）：この研究の目的は、消費者行動分析を想定した機械学習モデルの利用において、同一のデータおよび手法を用いても、同一の推定結果が再現されない問題について精査し、パラメータ推定の安定性を改善することである。この研究の成果は、主に次の2つにまとめられる。(1)制約を取り入れたLDAを推定するマルコフ連鎖モンテカルロ・アルゴリズムを提案し、推定値の安定性が改善されることを示した。(2)制約付きのニューラルネットワーク（オートエンコーダ）の推定によって、得られる結果の安定性を示した。しかし、これらは限定的な仮定での結果であり、より一般的な状況を想定した拡張が今後の課題である。

研究成果の学術的意義や社会的意義

本研究の学術的意義は、社会科学の領域で重視されている識別性の側面から機械学習の応用を議論した点であると考えられる。特にマーケティングや消費者行動分野において、消費者心理のような観測できない要因や不確かな要因を仮説的に測定したり、モデルで記述したりすることは重要な役割を果たしてきた。しかし、それらについて同一のデータと手法を用いても著しく異なる結果が得られる場合には、意思決定に多重の不確実性を与えることになる。このような文脈で本研究の社会的意義は、実社会で急速に利用が進んでいる機械学習について、社会科学での理論的背景も考慮しながらより頑健な応用方法を探索した点であると考えられる。

研究成果の概要（英文）：This study aims to investigate the identifiability of machine learning models such as latent dirichlet allocation (LDA) and neural network to improve the stability of their parameter estimations in consumer behavior analysis. The results of this study provide mainly two contributions. (1) LDA: The stability and identifiability of LDA in specific situations were clarified with simulation-based parameter recovery experiments. The author proposed Markov Chain Monte Carlo algorithms for the Bayesian estimation of constrained LDAs and showed their stability. (2) Neural network (autoencoder): The author proposed constrained estimation methods for neural networks and indicated that they provide stable and explainable results compared with an unconstrained standard neural network. However, these two main results were obtained under limited assumptions, so that the author will explore the extension and scope of the applications in the future study.

研究分野：マーケティング・リサーチ

キーワード：識別性 潜在的ディリクレ配分法 ニューラルネットワーク 潜在変数モデル マーケティング尺度  
消費者心理 消費者行動

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

## 1. 研究開始当初の背景

現代では、ビッグデータの利活用に向けた機械学習の開発や応用が著しく進んでいる。マーケティング分野では特に、高次元・非構造データの次元圧縮や潜在変数の推定を目的とする機械学習の多くが、製品や消費者の潜在特性を抽出する新しい方法として注目されている。一方、マーケティング分野には計量心理学から導入された心理特性の測定理論とその応用が従来から議論されてきている。これに従えば、理論や経験に基づく「構成概念」を操作的に仮定することによって、直接には観測できない消費者特性の測定を潜在変数の推定に置き換える方法がある。ここで構成概念とは、現実に観測することは困難であるが、一般に、もしくは理論的に、その存在が認められており、人間の行動規範や意思決定の基になっているものを表す。マーケティング分野であれば、顧客が知覚した満足度やサービス品質、また、購買経験や消費経験の中での様々な感情や認知などが含まれ、消費者行動の予測や消費者心理の理解に役立っている。

このような従来の方法論と比較すると、マーケティング分野では機械学習の応用が進む一方で、これまで重視されてきたパラメータや潜在変数の識別性の問題を具体的に議論している研究は少ない。パラメータが識別されないモデルとは一般に、モデルを特徴づけるパラメータの最適解が一意に定まらないモデルのことを表す。このとき、分析者や意思決定者にとって興味のあるパラメータの値や範囲に想定していなかった複数の候補が存在することになるため、これを適切に推定することはできないことを意味する。このようなモデルに基づく推論や意思決定には、予期せぬ多重の不確実性を伴うことになる。

## 2. 研究の目的

本研究の大きな目的は、従来の潜在変数モデルと関連の深い Latent Dirichlet Allocation (LDA) とニューラルネットワークに焦点をあて、これらのモデルで起こり得る識別性の問題やパラメータ推定の不安定性の解決に取り組むことである。また、これらのモデルを活用したより信頼性や頑健性の高い計量マーケティング・モデルを構築し、現実の消費者行動分析への応用を議論することである。さらに、本研究はマーケティングや消費者行動および心理に関連する構成概念の新しい測定方法についても検討することで、伝統的な手法の応用範囲の拡張を目指している。

## 3. 研究の方法

### (1) LDA について

LDA と関連のある非負値行列分解におけるアンカーワードの仮定や因子モデルの識別条件に基づき LDA に制約を取り入れ、その制約付き LDA を推定するためのマルコフ連鎖モンテカルロ・アルゴリズムの検証を行った。具体的には、シミュレーションベースのパラメータリカバリの実験により、制約のない LDA のパラメータ推定値の不安定性や、提案した方法の安定性について評価した。

### (2) ニューラルネットワークについて

ニューラルネットワークと関連のある因子モデルの識別条件を基に制約を取り入れた推定方法を検討した。具体的には、マーケティング尺度を利用したオンライン調査を実施してデータを収集し、提案されたニューラルネットワーク・モデルによる分析によって質問項目と潜在因子の理論的な関係性が再現されるかを検証した。

## 4. 研究成果

### (1) LDA について

個々のデータにおける合計頻度が少ないデータセットを分析する場合や、母集団または観測データに出現するトピックに偏りがある場合などの特定の状況において、通常の LDA によるパラメータの再現性は低いことが示された。また、各トピックの単語割合を表すパラメータの大小関係が再現されても、トピック割合が不安定になることがある。この結果は、実際のデータ分析において単語割合から各トピックの解釈が安定的に得られても、それらの潜在トピックを予測変数として何らかの目的変数を予測する回帰分析に用いると、推定される回帰係数の大小関係が不安定になる(例えば、真のデータ生成プロセスでは重要度の高いトピックについて、その重要度が低く推定される場合がある)ことを意味する。したがって、推定されたトピックを予測だけでなく、目的変数に対するトピックの関係性や影響力を比較するために用いる場合には、適切な解釈が得られないことがある。

一方で、提案手法により、ある1つのトピックのみに割り当てられる単語(アンカーワード)を固定したり、トピック割合を表すパラメータの事前分布を推定したりすることによって、パラ

メータの再現性が改善されることを示した。しかし、本研究で用いたシミュレーションの設定は、文書データというよりもマーケティング尺度データを想定したものになっている。そのため、データ生成過程の状況設定が限定的であり、今後の課題として、単語やトピックの次元を増やした場合などについて、より広範囲に精査する必要がある。

## (2) ニューラルネットワークについて

第1段階では、提案した手法によってニューラルネットワークのパラメータ推定値が、意図したように制約されるかを実証分析によって検証した。その結果、因子モデルで課されるいくつかの制約が満たされるようにパラメータや隠れユニット（潜在因子）が推定されることを示した。さらに、マーケティング尺度データの分析において、各潜在因子（構成概念）と理論的に関連する質問項目の少なくとも1つとの関係が重み行列の中で固定されるように制約することで、他の質問項目についても想定される理論的な関係性が再現されることを示した。

第2段階では、マーケティング尺度と文書データの同時分析を行うニューラルネットワーク・モデルを提案し、制約あり・なしのニューラルネットワークについて性能比較を行った。ここで、制約ありのニューラルネットワークには、潜在因子と項目得点間の重み行列に確認的因子分析と等価の制約を課している。その結果、データへの適合度を表す指標として用いた平均二乗誤差（RMSE）に大差はない一方で、制約ありのニューラルネットワークでは、パラメータ推定値の安定性は大幅に改善することが示された。この制約ありのニューラルネットワークを利用することで、文書データからマーケティング尺度に基づく因子の解釈が得られるメリットがある。さらに、通常はマーケターや分析者が必要とする項目得点を含むはずのない文書データから理論的に想定される消費者心理を測定、また可視化したり、その結果に基づき消費者行動を予測したりすることに役立つと考えられる。

## 5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Toshikuni Sato	4. 巻 -
2. 論文標題 Penalizing Neural Network and Autoencoder for the Analysis of Marketing Measurement Scales in Service Marketing Applications	5. 発行年 2021年
3. 雑誌名 AI and Analytics for Smart Cities and Service Systems. ICSS 2021. Lecture Notes in Operations Research	6. 最初と最後の頁 31~42
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-90275-9_3	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Toshikuni Sato	4. 巻 -
2. 論文標題 A Penalized Neural Network Model for Predicting Unobserved Scores of Construct Indicators and Reproducing Latent Scores of the Theoretical Constructs by Using Text Information	5. 発行年 2022年
3. 雑誌名 SSRN Electronic Journal	6. 最初と最後の頁 1~10
掲載論文のDOI（デジタルオブジェクト識別子） 10.2139/ssrn.4058970	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Toshikuni Sato	4. 巻 -
2. 論文標題 Bayesian Estimation for Identifiable Topic Models with Latent Dirichlet Allocation	5. 発行年 2020年
3. 雑誌名 SSRN Electronic Journal	6. 最初と最後の頁 1-24
掲載論文のDOI（デジタルオブジェクト識別子） 10.2139/ssrn.3721769	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計6件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 Toshikuni Sato
2. 発表標題 Identification of Latent Dirichlet Allocation for Theoretical Construct Measurements with Marketing Scales
3. 学会等名 2021 Frontiers in Service Conference（国際学会）
4. 発表年 2021年

1. 発表者名 Toshikuni Sato
2. 発表標題 Penalizing Neural Network and Autoencoder for the Analysis of Marketing Measurement Scales in Service Marketing Applications
3. 学会等名 2021 INFORMS Virtual Conference on Service Science (国際学会)
4. 発表年 2021年

1. 発表者名 佐藤平国
2. 発表標題 マーケティング尺度を用いた測定モデルに対する罰則化法 - 消費者心理の識別と予測への応用 -
3. 学会等名 第63回消費者行動研究コンファレンス
4. 発表年 2021年

1. 発表者名 佐藤平国
2. 発表標題 ベイズモデルの識別性の再考とマーケティング尺度の応用可能性
3. 学会等名 ID-POSデータのマーケティングの活用研究部会
4. 発表年 2021年

1. 発表者名 佐藤平国
2. 発表標題 識別可能なLDAトピックモデルのベイズ推定法
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 佐藤平国
2. 発表標題 消費者行動研究のための識別可能なLDAトピックモデル
3. 学会等名 第61回消費者行動研究コンファレンス（消費者行動研究学会）
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------