

令和 4 年 6 月 14 日現在

機関番号：34310

研究種目：若手研究

研究期間：2019～2021

課題番号：19K20226

研究課題名（和文）解釈容易性を担保した次元縮約法の開発

研究課題名（英文）Dimensional reduction method with interpretable estimated values

研究代表者

土田 潤 (Tsuchida, Jun)

同志社大学・文化情報学部・助教

研究者番号：40828365

交付決定額（研究期間全体）：（直接経費） 1,800,000円

研究成果の概要（和文）：本研究では、解釈容易性の指標としてGini Indexを利用し、Gini Indexが一定上という制約条件を課した次元縮約法の開発を行った。Gini Indexで罰則を付けることで、得られる推定値の一部が0となるだけでなく、推定値間の差が大きく、メリハリがついた値を得ることができる。成果として、主成分分析にてGini Indexが一定上という制約条件を課した手法とそのパラメータ推定法の開発、因子分析における回転法の開発の2つの手法を開発した。成果は、学会発表を通して世間に公表した。

研究成果の学術的意義や社会的意義

“解釈容易性”の議論は実用上重要であるが、現在、積極的に議論されていない。本研究では、Thurston(1947)が言葉でのみ定義した解釈容易性を、式によって表現することを目標とした。解釈容易性を式によって表現することで、実用上重要な問題である“解釈容易性”を担保した新しい統計手法の構築の基礎を作ることができる。本研究ではGini Indexを用いて解釈容易性が定義できるかを検証し、解釈容易性を最大化する次元縮約法を開発した。本研究の成果は、新しい統計手法の構築の基礎の一助となる。

研究成果の概要（英文）：In this study, we used the Gini Index to measure the interpretability of estimated values. We developed dimensional reduction methods with the constraint that the Gini Index must be above a particular value. From this constraint, we could obtain the sparse estimated values with a large variance. This characteristic of the estimated values corresponds to the interpretability.

We have developed two methods: One is principal component analysis with the constraint that the Gini Index must be above a certain level. The other is a rotation method for factor analysis. We reported the study results to the public through conference presentations.

研究分野：統計科学

キーワード：主成分分析 因子分析 因子回転法 交互最小2乗法

## 1. 研究開始当初の背景

リアルデータの活用では、大規模性を生かすために、深層学習を用いた方法やアンサンブル学習を用いた方法等が適用され、優れた予測精度を持つことが報告されている。しかし、リアルデータに求められることは予測ばかりではない。ある問題について、リアルデータを用いて解決することを考えると、最終的には得られた知見を社会還元することが求められる。現象を正しく理解し、最終的に得られた知見を社会還元することがリアルデータを活用する上で求められている。予測を含むデータ利用を適切に行うためには、その前段階であるデータ理解のフェーズが重要である。データ理解を深めるためには、基礎集計と適切な変量の作成が必要となる。作成した新たな変量が問題に対して適切であるかを検証し、最終的に必要な変量を決定するまでが、データ理解である。本研究では、データ理解のフェーズにおける問題を扱う。

データ理解を深めるための有用な方法の1つである次元縮約法に着目する。データ理解のために次元縮約法を適用する場合、得られる新たな変量がどのようなものかを説明できなければならない。もし説明できなければ、データ理解のフェーズにもかかわらず、得られた変量がブラックボックスになる。しかし、実際は、膨大な数の変量から新たな変量が作成されるため、得られた変量の解釈が困難である場合が多い。そのため、解釈容易な次元縮約法が実用上求められている。

得られた変量の解釈を容易にするため、多くの研究者によってスパース推定を伴う次元縮約法が提案されている(Zou, et al., 2006; Shen and Huang, 2007; Adachi and Trendafilov, 2016)。これらの手法の多くは、「いくつかのパラメータの推定値が0であればよい」という指標をもとに作成されている。いくつかのパラメータの推定値を0にすることで、必要な変数を選択することができるためである。しかし、変数を選択することは解釈容易の必要条件であり十分条件ではない。

次元縮約法によって得られる新たな変量が解釈容易であるかどうかの条件についてはThurston(1947)によって議論されている。Thurston(1947)は、新たな変量の構造が単純であるための条件を5つ提示している。しかし、その条件は言葉でのみで定義され、定式化に至っていない。解釈容易性は5つの条件のうち2つに焦点が置かれて議論されている。

その1つが「いくつかのパラメータの推定値が0とであればよい」という指標であり、スパース推定として議論されている。もう1つは「パラメータの推定値が互いに大きく異なる」という指標であり、分散によって定式化した方法が存在する(Kaiser, 1958; Neuhaus and Wrigley, 1954)。解釈容易性の議論はこの2つの立場によって独立して発展しているが、これらを統合した解釈容易性の定式化は十分に検討されていない。さらに、解釈容易性を担保した次元縮約法についても、2つの立場の一方のみを取り上げた議論に限られている。

## 2. 研究の目的

本研究の目的は解釈容易性を定式化し、新たな解釈容易な次元縮約法を開発することである。先に述べたように、次元縮約法では(1)スパース推定を伴う次元縮約法、(2)推定値間の分散を大きくする次元縮約法の2つが主に研究されている。(1)、(2)の研究分野は独立に発展しており、これらを統合されている研究は少ない。

本研究では、これら2つの研究分野の長所を組み合わせ、Thurston(1947)が示した解釈容易性の条件を達成する。つまり、いくつかの推定値を0とし、かつ、0でない推定値の大小関係を明確にする関数を定式化し、そして、定式化された解釈容易性を制約条件とした次元縮約法を開発する。これにより、データの理解のフェーズに貢献する。次元縮約法には主成分分析を対象として開発を行う。因子分析の回転法も対象とする。

## 3. 研究の方法

本研究では、不平等指数として用いられているGini Indexに着目し、解釈容易性の定量化を試みる。Gini Indexは(1)、(2)の2つの組み合わせの定式化の1つとして解釈できることが知られている。しかし、Gini Indexは推定値の大小関係を明確にしやすいが、いくつかの推定値を0としにくい。そのため、変数の数が膨大にあるとき、解釈容易性の指標としては十分ではない可能性がある。本研究ではいくつかの推定値を0としやすくするために、ミンコフスキーノルムでGini Indexを考察し、その特徴を調査する。

さらに、ミンコフスキーノルムGini Indexが一定以上あることを制約条件とした次元縮約法について考え、データの理解のフェーズに用いられる統計的手法を開発する。次元縮約法の開発にはその代表手法である主成分分析を軸に開発する。主成分分析は様々な次元縮約法の基礎と

なっており、この開発をきっかけに、多くの手法への拡張の基礎を作ることができる。さらに、因子分析法での負荷量に対する回転法についても Gini Index を利用できるか検証する。因子分析法での負荷量に対する回転法は次元縮約法の中でも初期に解釈容易性について議論されている。因子回転法の文脈で、ミンコフスキーノルム Gini Index 最大化と既存法との関係性を明らかにすることによって、ミンコフスキーノルム Gini Index の解釈容易性の指標としての立ち位置を明らかにする。

#### 4. 研究成果

本研究では次の2つの方法を開発した。

主成分分析のパラメータを解釈容易に推定する方法を開発した。具体的には、パラメータ推定の目的関数に Gini Index 罰則を付与し、同時に最適化することで、いくつかを0として推定し、かつ、0でない推定値の大小関係を明確に推定する方法である。既存の罰則項を付与方法とは異なり、Gini Index は非凸関数であるため、最適化が困難である。そこで、Gini Index の Majorization-function を導出し、凸最適化問題にパラメータ推定の問題を書き直した。これにより、単調減少列を作ることができ、局所解を効率的に得られるアルゴリズムを構築した。

このアルゴリズムは、主成分分析のみに利用できるわけではなく、線形回帰分析にも適用可能である。実際、提案アルゴリズムは主成分分析を回帰分析の問題に書き直した方法をもとにして構築されている。したがって、提案アルゴリズムは回帰分析にも適用可能だけでなく、回帰分析と更新式が同等になるようなアルゴリズムに対して適用可能となる。特に次元縮約法においては、Tukcer-3 と呼ばれる主成分分析の拡張法が該当する。Tukcer-3 は主成分分析だけでなく多くの多変量解析手法の拡張となっているため、提案アルゴリズムを用いることで、多くの多変量解析手法に解釈容易な解を求めるアルゴリズムを提供できる可能性がある。

提案法を簡単なデータに適用したところ、0でない推定値の大小関係を明確に推定することが可能となった。しかし、他のスパース推定法に比べ、0と推定する個数がやや少ない傾向にあった。よりスパース性を高めるために他の方法と組み合わせることは今後の課題である。

因子分析の回転法として Gini Index を最大にするような回転法を開発した。因子分析の回転法は大きく分けて直交回転と斜交回転の2つに分類できる。本研究では直交回転と斜交回転の2つの手法の両方を開発した。具体的には、それぞれアルゴリズムは導出した Gini Index の Majorization-function をもとにしている。この成果である Majorization-function をさらに、微分可能なものを書き直すことで、既存のプロクラステス回転の問題に落とし、その結果として、直交回転と斜交回転の2つの手法の両方を開発した。

既存法との関係について考察も行った。Gini Index をユークリッド距離の2乗の意味で確認すると、分散最大化に対応する。これは、Kaiser (1958)の方法と等価な方法である。Gini Index 最大化問題は Kaiser (1958)の方法の拡張と考えることができる。Kaiser (1958)の方法は古典的な因子回転法であるが、非常に多く利用されている回転法である。さらに、多くの手法が、Kaiser (1958)の方法をもとに提案されている。

Gini Index を最大にするような回転法については、一部恣意性が残る。これは、Gini Index を計測するベクトルが、行ベクトルか列ベクトルかを選択しなければならないためである。開発した手法は、これらの選択を行ベクトルで計算された Gini Index と列ベクトルで計算された Gini Index の重み付き和として提案しているが、この重みは恣意的に決められる。Gini Index の最大値が次元に依存するため、提案法では、次元数に応じた重みを付与した。解釈容易性のためには、いくつかの重みパラメータを試し、実行する必要がある。これらの方法については今後の課題である。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計6件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 土田潤, 宿久洋
2. 発表標題 Gini係数を用いた斜交回転法について
3. 学会等名 日本分類学会第39回大会
4. 発表年 2020年

1. 発表者名 Jun Tsuchida, Hiroshi Yadohisa
2. 発表標題 Principal component analysis using the Gini coefficient penalty function
3. 学会等名 Data Science, Statistics and Visualization 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 土田潤, 宿久洋
2. 発表標題 Gini 係数をペナルティ関数とした主成分分析について
3. 学会等名 応用統計学会2019 年年会
4. 発表年 2019年

1. 発表者名 土田潤, 宿久洋
2. 発表標題 Gini 係数を目的関数とした直交回転法について
3. 学会等名 日本分類学会第38 回大会
4. 発表年 2019年

1. 発表者名 土田潤, 宿久洋
2. 発表標題 A majorization-minimization algorithm for Gini coefficients penalized regression
3. 学会等名 2021年度統計関連学会連合大会 (国際学会)
4. 発表年 2021年

1. 発表者名 Jun Tsuchida, Hiroshi Yadohisa
2. 発表標題 A rotation method by using the Gini coefficient
3. 学会等名 Joint Statistical Meeting 2021 (国際学会)
4. 発表年 2021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