

博士学位論文審査要旨

2022年5月31日

論文題目： Complex Linear Regression Based on Structural Latent Variable Modeling (構造的潜在変数モデリングに基づく複雑な線形回帰)

学位申請者： Doung Thi Binh An

審査委員：

主査：	文化情報学研究科	教授	宿久	洋
副査：	文化情報学研究科	特別客員教授	吉野	諒三
副査：	文化情報学研究科	教授	金	明哲
副査：	文化情報学研究科	教授	鄭	躍軍
副査：	大阪大学大学院人間科学研究科	教授	足立	浩平

要旨：

本論文は、顕在変数間の複雑な関係性を明らかにするための統計的分析法に関するものである。モデルに潜在変数を導入し、潜在変数と顕在変数間の線形関係を仮定することにより、潜在変数の解釈容易性を高めることを提案している。さらに、潜在変数を用いた柔軟なモデリングにより、顕在変数間の複雑な関係も表現できることが特徴である。これまで、複数の潜在変数を用いたモデリングは提案されていたが、パラメータ推定のための制約が強く、柔軟なモデリングができていたとは言い難かった。本論文では、パラメータ推定法について2つの方法を提案しており、提案法により、パラメータの制約が緩和され、モデリングの幅が広がったと考えられる。

本論文は5章から構成されている。第1章では本研究の背景・目的・構成、第2章では研究に使われる表記法について説明している。第3章では、提案法である多変量直交回帰分析のモデルとその推定法、第4章では、構造方程式モデリングとその推定法である一般化最大エントロピー法に関する方法について述べ、一般化最大エントロピー法の制約式をK-means法で緩和した方法について扱っている。第5章では、本論文での方法論としての貢献と実データ分析にどのように活かせるかを論じている。

本論文での主な貢献は、解釈が容易でありながらも複雑な関係をモデリングする統計的方法に関して、2つの観点から新たな方法を提案している点とボトルネックであったパラメータ推定の制約を緩和した点の2点である。よって本論文は、博士(文化情報学)(同志社大学)の学位論文として十分な価値を有するものと認められる。

総合試験結果の要旨

2022年5月31日

論文題目： Complex Linear Regression Based on Structural Latent Variable Modeling (構造的潜在変数モデリングに基づく複雑な線形回帰)

学位申請者： Doung Thi Binh An

審査委員：

主査：	文化情報学研究科	教授	宿久	洋
副査：	文化情報学研究科	特別客員教授	吉野	諒三
副査：	文化情報学研究科	教授	金	明哲
副査：	文化情報学研究科	教授	鄭	躍軍
副査：	大阪大学大学院人間科学研究科	教授	足立	浩平

要旨：

学位申請者は2016年4月より本学大学院文化情報学研究科博士後期課程に在学し、2020年3月に単位取得後退学したのち、2022年4月より再入学している。この間、国内会議および国際会議での研究発表を通じて研究活動を積極的に行い、それらの成果を論文誌に3本公刊している。また、英語の語学試験にも合格していることから語学（英語）について十分な能力を有していると認定されている。

申請者の学位申請に関し、2022年5月27日（金）18:30から約1時間の公聴会と20分の審査会において、種々の質疑応答を行った。申請者は研究内容及び関連する質問に対する確に対応したことで、学位論文審査委員会は申請者が博士（文化情報学）（同志社大学）の学位を授与するに十分な学力を有することを確認した。

よって、総合試験の結果は合格であると認める。

博士學位論文要旨

論文題目： Complex Linear Regression Based on Structural Latent Variable Modeling

(構造的潜在変数モデリングに基づく複雑な線形回帰)

氏名： Duong Thi Binh An

要旨：

Models are created in order to simulate and solve real-world problems. Linear models have a similar purpose. The more a model describes the actual condition, the more effective it is as a reference while making decisions. Linear models are widely applied in practice and quantitative research fields because they are easy to understand and interpret. However, problems in practice or research are complicated, requiring these models to be modified; for example, the number of variables is increased compared with the original model, assumptions are added to the models, or constraints are replaced. While limitations on statistical techniques, such as a lack of follow-up models to solve problems, or theories exist, there is no fixed method that can be applied to modified cases, or the estimation method exists but has many restrictions that make it difficult to apply.

Within the scope of this dissertation, we aim to make two key contributions to the multivariate linear model literature with these complex models expressed through structural latent variables:

- In Chapter 3, we propose an estimation method for a linear model containing numerous independent and dependent variables that all have errors. It can also be seen as a modification of a multivariate linear regression, wherein the number of dependent variables is increased and assumptions pertaining to errors are added to the model.
- In Chapter 4, we focus on a more complex model, namely, structural equation modeling (SEM). Essentially, SEM is a set of multivariate linear regressions wherein the dependent variable in one equation can be an independent variable in another equation and vice versa. The estimation method proposed in this chapter is a constraint improvement for the generalized maximum entropy (GME) for SEM.

Before going to key chapters, a general context is also provided in Chapter 1. Finally, we present the conclusion to summarize the major contributions of the proposed methods in Chapter 5.

In Chapter 3, we propose a multivariate multiple orthogonal linear regression (MMOLR). The MMOLR expresses the relationship between two sets of dependent and independent variables. The MMOLR especially considers the advantages of the errors-in-variables (EIV) model, that is, the errors are included in all independent and dependent variables. Consequently, the assumptions of the model are easy to satisfy in practice. Our contribution is deriving an estimation method. It is in this context of total least squares that we reveal the relationship between the MMOLR and the canonical regression model.

Next, in Chapter 4, we derive a novel estimation method for SEM. SEM is widely used in many fields such as psychology, behavioral science, and marketing, to measure unobservable concepts and explore complicated relationships. Over the past four decades, especially, the maximum likelihood (ML) has become the predominant estimation method for SEM. However, this method relies upon sample size to ensure a normal distribution for data and non-negative degree of freedom: The more complicated the model is, the larger the sample size that is required. This is particularly evident in the present world, wherein researchers must consider increasingly multi-parameter problems. This increase in concepts gives rise to an upsurge in relationship paths. At the same time, research subject matters and/or populations with sensitive issues (e.g., managers of enterprises or patients with social diseases) present challenging barriers to data collection. Therefore, an estimation method that does not require assumptions on data distribution but works with minimum sample sizes is necessary. We contribute to this by proposing a new estimation method, the K-means GME for SEM. The K-means GME-SEM can reduce the volume of calculation and redundant constraints through the K-means centroid as a representative for data in consistency constraints. The simulation results also confirm that the new method improves the appropriate goodness of fit. Although there are many limitations to the proposed method, the results of this study are a major step toward a new approach for an information-theoretic-based SEM.