

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 15 日現在

機関番号：12102

研究種目：研究活動スタート支援

研究期間：2014～2015

課題番号：26885007

研究課題名(和文) 縦断データの分類手法に関する方法論的研究と教育・発達心理学研究への応用

研究課題名(英文) Methodology for classification of subjects through longitudinal data and its applications for educational and developmental psychology

研究代表者

宇佐美 慧 (USAMI, Satoshi)

筑波大学・人間系・准教授

研究者番号：20735394

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：縦断データに見られる変化のパタンの個人差が、個人のどのようなプロフィール(例えば、性別や年齢)により説明・分類できるのかを調べる上でSEMTrees(構造方程式モデル決定木)と呼ばれる方法が近年注目されている。SEMTreesでは、変化の様態を記述するテンプレートモデルを立ててデータを分割していく。本研究課題では、サンプルサイズの大きさやモデルの誤設定の有無など多要因を操作しながら、大規模シミュレーションに基づくSEMTreesの性能評価と高速化の研究に取り組んだ。また、正確な分類に必要なサンプルサイズに関する方法論的検討や、今後利用が見込まれる種々のテンプレートモデル間の数理的比較も行った。

研究成果の概要(英文)：When conducting longitudinal research, investigation of between-individual differences in patterns of within-individual change can provide important insights into a variety of typical developmental and growth patterns. We focused on a model-based exploratory data mining technique called structural equation model trees (SEMTrees). SEMTrees is a technique that classifies participants into subsets (classes) by recursively partitioning data based on the values of the observed covariates such as age, sex and personality of individuals. In the present research we investigated the performances of SEMTrees under various research designs including an influence of possible model misspecifications, and also investigated a relation between the sample size and precision of classification under various template models specified.

研究分野：心理統計学

キーワード：縦断データ 決定木 構造方程式モデル 分類 ソフトウェア モデルの誤設定

1. 研究開始当初の背景

小学生を対象に毎年実施して得た身体測定データ(図1上)のように、複数の個人から継続的に測定されて得たデータを縦断データ(longitudinal data)と呼ぶ。縦断データでは同じ個人を追跡するため、たとえば成長に伴って身長がどのように伸びていくかといった測定値の変化のパターンや、またその変化のパターンの個人差を吟味できる。このような利点から、教育・発達心理学の領域を中心とした縦断データの収集、および縦断データの分析手法の開発研究が世界各地で進んでいる。

実際の縦断データでは、成長に伴って身長が毎年概ね均等(=直線的に)に伸びる児童や、ある特定の学年を境に急速に伸びる児童もいるように、変化のパターンは個人に応じて異なる場合が多い。このような変化の多様性があるとき、特に数十~数千の独立変数を含む大規模な縦断調査において、「縦断データの変化のパターンの個人差が個人のどのようなプロフィール(性別・年齢・職業などのデモグラフィック変数、パーソナリティ特性などの心理学的変数、世帯年収や社会関係性など経済・社会的変数: 図1上では例として、性別、部活動の有無といった変数を挙げている)によって効率的に分類・説明できるのか」という問題は、発達の個人差の説明要因を理解する上で重要なだけでなく、(独立)変数の数が非常に多いビックデータの解析手法としても強い関心もたれる。ところが、この問題は、多様な縦断データの変化のパターンを推定しながらその個人差を説明する独立変数を同時に探索し分類する必要があるため、技術的水準の高い難題である。

この問題を解決する上で注目を集めている一つの方法が、Brandmaier, Oertzen, McArdle & Lindenberger (2013)により Psychological Method誌に発表された、構造方程式モデル決定木(Structural Equation Model Trees, 以下, SEMTrees)である。SEMTreesでは、直線的な変化や二次曲線的な変化など、縦断データの変化のパターンを表現するモデル(テンプレートモデル)を設定して、各ノード(小集団)内の変化のパターンを推定しながら、その個人差(集団差)を説明する独立変数を探索し、順々に対象全体を分割していく。例えば図1下では、直線的な身長の変化を表現するテンプレートモデルを設定したときに、数ある独立変数(図1上)のうち、性別が身長の変化の個人差を最もよく説明する変数と推定され、次に男児(という小集団)では



図1: SEMTrees のイメージ図

さらに部活動の有無で説明できると推定された。これより、身長の変化の個人差を説明する分類結果として、変化の傾きや切片(小1の値)が異なる計3つのノード(小集団)が抽出された(図1下)。

2. 研究の目的

SEMTreesは縦断データ解析研究におけるブレイクスルーとして、国内外で大きな注目を集めている。しかし、Brandmaier et al (2013)の論文は、SEMTreesの導入的な説明を主目的として書かれており、その方法論自体の適切性やありうる問題点を示す為の理論的検証はまだ十分なされておらず、またSEMTreesの応用事例数も現段階では限られている。そこで本研究ではSEMTreesに関する課題として、

- (1)大規模シミュレーションに基づくSEMTreesの性能評価研究
- (2)SEMTreesの実行の高速化の研究
- (3)大規模縦断データを用いたSEMTreesの応用研究

(4)テンプレートモデルの設定方法やサンプルサイズ(=個人の数)の違いが、分類精度や分類結果に与える影響についての基礎的研究

の4点の遂行を目的とした。

3. 研究の方法

1) 大規模シミュレーションに基づく SEMTrees の性能評価研究.

サンプルサイズ(=個人の数)や測定時点数およびテンプレートモデルの違い等に応じて、実際のカテゴリ結果が真のカテゴリ(グループ)構造をどの位の精度で再現できるかについての詳細は、まだ明確にはわかっていない。また、変化のパターンを説明する真のモデルとの乖離が大きいテンプレートモデルを用いた場合、分類結果がどの程度歪むのかというモデルの誤設定(model misspecification)の影響も十分検討されていない。このように、強い応用上のニーズに反し SEMTrees の理論的整備は現段階では十分でない。そこでこれらの問題を同時に網羅的に検証するため、大規模シミュレーションを行い、SEMTreesによる分類の再現精度の説明要因を調べた。

2) SEMTrees の実行の高速化の研究.

Brandmaier et al (2013)では、フリーの統計ソフト R を用いて SEMTrees を手軽に実行するためのパッケージを提供している。しかし、SEMTrees の実行にはしばしば膨大な計算量を必要とするため、必ずしも現実的な時間内に計算が終わらないという実用上重大な問題がある。そこで、特に連続的な独立変数を多く扱う場合に着目して計算アルゴリズムを改良し、SEMTrees の実行の高速化を図った。

3) 大規模縦断データを用いた SEMTrees の応用研究

全国高齢者縦断調査や中高生の精神保健調査を用いた SEMTrees の応用事例を報告する。これらを通して、例えば前者のデータでは、高齢者の身体的・精神的健康の水準が加齢とともにどのように変化しているのか、またその変化のパターンの個人差は、高齢者の保有する様々な資源や生活状況、および社会関係などの違いからどのように説明・分類できるのかを調べる。

4) テンプレートモデルの設定方法やサンプルサイズの違いが、分類精度や分類結果に与える影響についての基礎的研究.

設定したテンプレートモデルの違いやサンプルサイズの違いが分類精度や母数の推定精度などの結果にどのように影響を与えるのかを調べるための基礎的検討を行った。

具体的には、SEMTrees の実際の応用でよく利用されると考えられる、潜在成長モデルなどの個別の統計モデルやその他の分類のための統計モデルにおける、一定以上の母数の推定精度および分類精度、検定力の確保に必要なサンプルサイズおよびモデルの設定方法についての基礎的検討を行った。さらに、SEMTrees での実際の応用が期待される、個別の統計モデル間の数理的な関係性についての基礎的な検討も行った。

4. 研究成果

(1)について、他の既存のカテゴリ手法との性能比較およびモデルの誤設定の影響を踏まえた大規模なシミュレーションを実施し、SEMTrees による分類精度やその規定因に関する知見が網羅的に得られた。その成果をまとめ、現在海外誌に投稿中である。モデルの誤設定の影響については、SEMTrees の解説論文においても扱った。(2)については、推定の高速度とともに、新たな分類基準に基づく SEMTrees の分析プログラムを開発した。現在、(3)の縦断データへの適用事例とともに、プログラムの概要や分類のアルゴリズムを解説した論文を執筆中である。(4)に関しては、縦断データ分析の為に様々な統計モデルにおけるモデル設定とクラス(小集団・グループ)数の推定精度の関係や、サンプルサイズ決定方法の開発、および縦断データ分析の為に利用される様々なテンプレートモデル間の数理的関係性などの基礎的研究を進めた。これらの成果の一部は既に海外誌に掲載されており、その他は現在海外誌に投稿中である。

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 5 件)

1, 宇佐美慧 (2016). 構造方程式モデル決定木 (SEMTrees) の方法と課題—モデルの誤設定について— 計算機統計学 印刷中 査読有

2, Todo, N., & Usami, S. (2016). Fitting finite mixture models in longitudinal design: A recommendation for model selection and estimation of the number of classes. Structural Equation Modeling. in press. 査読有

3, Usami, S. (2016). Generalized sample size determination formulas for evaluating contextual effects in three-level data. Psychometrika. in press. 査読有

4, Usami, S., Hayes, T., & McArdle, J. J. (2015). Inferring longitudinal relationships between variables: model selection between the latent

change score and autoregressive cross-lagged factor models. *Structural Equation Modeling*, 23, 331-342. 査読有

5, Usami, S., Hayes, T., & McArdle, J.J. (2015). On the mathematical relationship between latent change score model and autoregressive cross-lagged factor approaches: Cautions for inferring causal relationship between variables. *Multivariate Behavioral Research*, 50, 676-687. 査読有

[学会発表] (計 2 件)

1, Usami, S., Hayes, T., & McArdle, J.J. (2015). Some notes on the mathematical relationship between two longitudinal models for assessing causality. 行動計量学会第43回大会 首都大学東京 東京都八王子市 9月3日

2, Usami, S., Hayes, T., McArdle, J.J. (2015). Inferring longitudinal relationships between variables: Model selection between the latent change score and autoregressive cross-lagged factor models. Paper presented at the 80th Annual Meeting of Psychometric Society. Beijing Normal University, China, July 13th.

[その他]

ホームページ等

<http://www.satoshiusami.com/>

6. 研究組織

(1) 研究代表者

宇佐美 慧 (USAMI, Satoshi)
筑波大学・人間系・准教授

研究者番号: 20735394