

直交制約に基づく非計量非負値行列因子分解について

同志社大学大学院 谷岡 健資

同志社大学 宿久 洋

1. はじめに 質的多変量データは社会調査等様々な分野で観測され、対象とカテゴリ間や変量間の関係を把握することは重要となる。そのような目的を達成する方法として、多重対応分析や非計量主成分分析 (Murakami et al., 1999) が挙げられる。多重対応分析の利点は対象とカテゴリ間の関係を視覚的に把握することができる点であり、非計量主成分分析では対象と変量間の関係を低次元空間で把握することができる点であると考えられる。しかし、近年情報技術の発達に伴い、データが大規模複雑化し、既存手法では対象とカテゴリ間の関係や対象と変量間の関係を把握することが困難な場合がある。そのような問題を克服する方法として、本報告では非計量主成分分析のモデルを利用した直交制約に基づく非計量非負値行列因子分解法を提案する。

本提案手法は非負値行列因子分解法 (Lee and Seung, 1999) の考え方をを用いる。非負値行列因子分解法とは与えられた非負値データを低ランクの非負値行列で近似する方法のことを指す。非負値行列因子分解の利点は3点挙げられ、1点目は解釈すべき係数行列が非負値であることから符号の影響を考慮せずに解釈可能な点であり、2点目は非負値制約を課すことで各値がスパースとなる点であり、3点目は補助関数法によって容易に各パラメータを推定できる点である。質的多変量データはダミー変数行列と呼ばれる1,0の値をとりうるデータ行列によって表現されるため、非負値行列とみなすことができる。Li (2005) は binary matrix factorization と呼ばれる、1,0の値をとる低ランク行列で binary データを近似する方法を提案しているが、変量の要因を考慮できない点や低ランク行列を解釈する際に非常に単純な解釈しかできない点等の問題点が挙げられる。

そこで、提案手法では、binary matrix facotrization のような表現方法ではなく非計量主成分分析で採用されている各カテゴリに付与される数量化行列および各変量を評価する係数行列を用いて低次元空間を表現する方法を採用する。提案手法の利点としては非負値行列因子分解の利点に加えて、非負値かつ直交制約を課すことで係数行列が Perfect Simple Strcuture (Bernaad and Jennrich, 2003) となり、解釈が容易となる点である。Ding et al. (2005) は直交制約に基づく非負値行列因子分解法を提案しているが、初期値として係数行列の0と推定する要素をあらかじめ推定しなくてはならないという問題点がある。そのため、その問題点を克服するため、本手法では Adachi and Trendafilov (2013) を基に推定を行う。

2. 提案手法 提案手法の詳細および数値例については当日報告する。

参考文献

- [1] Adachi, K., and Trendafilov, N. T. (2013), "Perfect Simple Structure Factor Analysis with Alternating Least Squares Algorithm," 日本行動計量学会大会発表論文抄録集, **41**, 200–203.
- [2] Bernaad, C. A., and Jennrich, R. I., (2003), "Orthomax rotation and perfect simple structure," *Psychometrika*, **68**, 585–588.
- [3] Ding, C., He, X., and Ssimon, H. (2005), "Orthogonal nonnegative matrix tri-factorizations for clustering," *Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD)*, 126–135.
- [4] Lee, D. D., and Seung, H. S. (1999), "Learning the parts of objects with nonnegative matrix factorization," *Nature*, **401**, 788–791.
- [5] Li, T. (2005), "A general model for clustering binary data," *In SIGKDD*, 188–197.
- [6] Murakami, T., Kiers, H. A. L., & teb Berge, J. M. F. (1999), "Non-metric principal component analysis for categorical variables with multiple quantifications," Unpublished munuscript.