

Applying Non-negative Matrix Factorization with Covariates to the Longitudinal Data as Growth Curve Model

佐藤健一*

滋賀大学データサイエンス学部

非負値行列因子分解(Non-negative Matrix Factorization, NMF)^[1]を適用すると、非負の観測行列は同じく非負の基底行列と係数行列の積で近似できる。さらに、係数ベクトルが各個体の共変量で説明される場合にはパラメータ行列と共変量行列の積でかけ、tri-NMF^[2]の枠組みで記述することができる。この共変量を持つNMFは形式的には成長曲線モデル(GCM)^[3]の平均構造と等しい。GCMの基底行列は分析者によって与えられるのに対し、共変量を持つNMFの基底行列は未知であり、最適化されるという違いがある。本研究では共変量を持つNMFを経時測定データに適用し、GCMとの違いを比較する。また、共変量としてガウスカーネル関数を利用することで係数行列の高い精度での説明が可能になり、基底行列を介して観測値の予測も可能になる^[4]。さらに、この手法を実装したRパッケージ^[5]を公開した。時系列データ、テキストデータ、あるいは画像データなどの解析例を通し、その使い方や有用性を示す。特に、共変量としてガウスカーネル関数を利用することで係数行列を高い精度で説明でき、基底行列を介して観測値の予測も可能になった^[5]。

References

- [1] Lee DD, Seung HS: Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791. (1999) <https://doi.org/10.1038/44565>
- [2] Ding C, Li T, Peng W, et al: Orthogonal nonnegative matrix t-factorizations for clustering. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and datamining, pp126–135 (2006) <https://doi.org/10.1145/1150402.1150420>
- [3] Potthoff RF and Roy SN: A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika, 51, 313-326. (1964) <https://doi.org/10.2307/2334137>
- [4] Satoh K: Applying Non-negative Matrix Factorization with Covariates to the Longitudinal Data as Growth Curve Model. (2024) <https://doi.org/10.48550/arXiv.2403.05359>
- [5] R package: nmfkc for Non-Negative Matrix Factorization with Kernel Covariates <https://github.com/ksatohds/nmfkc>

*連絡先

〒522-8522 滋賀県彦根市馬場 1 丁目 1-1

0749-27-1067 内線(1067)

kenichi-satoh@biwako.shiga-u.ac.jp

