 Web版 ISSN 2189-0803 DVD版 ISSN 2189-079X

25pM1-9

テンソル分解を用いた教授無し学習による変数選択法のバイオインフォマティクスへの応用

中央大学　物理学科　田口善弘

Application of tensor decomposition basd unsupervised feature extraction to bioinformatics

Department of Physics, Chuo University, **Y-h. Taguchi**

筆者らはここ10年来、「主成分分析及びテンソル分解を用いた教師無し学習による変数選択法」を提唱し、バイオインフォマティクスの分野に適用してきた[1]。この方法は典型的なlarge p small n問題であるゲノムサイエンスにおける変数選択問題（例：二群で有意に差がある変数の選択）において有効に機能することが知られていたが、1）変数の選択数が少なく、偽陰性がたくさんあることが疑われた　2）いわゆるP値の頻度分布が帰無仮説（主成分または特異値ベクトルが正規分布する）と必ずしも整合的（P値の頻度分布は帰無仮説に従うならP値の頻度分布は一様分布）ではない　という問題があった。今回、この問題は帰無仮説に用いる正規分布の標準偏差を外れ値を除外して推定することで大きく改善することが分かった。標準偏差を最適化した提案手法を遺伝子発現プロファイル[2]、DNAメチル化[3]、ヒストン修飾[4]の各プロファイルに適用したところ、各プロファイルに適した変更を加えることなく、いずれもSOTAを凌ぐ性能が発揮できることが分かった。通常、プロファイルの種類ごとに値が異なった頻度分に従うため、個々に別種のアプリケーションが必要であり、同じプログラムで複数種のプロファイルの選択を行えることは稀である。本講演では主成分分析やテンソル分解を使うことで主成分や特異値ベクトルを考慮することになり、これらは元の値の線形結合なので乱数である場合には中心極限定理により、ガウス分布が出現することで、複数種のプロファイルを単一のプログラムで扱えるようになったらしいことなどを報告する。

References

[1] Y-h. Taguchi, Unsupervised Feature Extraction Applied to Bioinformatics
A PCA Based and TD Based Approach, Springer International, 2020.
[2] Taguchi, Yh., Turki, T. Adapted tensor decomposition and PCA based unsupervised feature extraction select more biologically reasonable differentially expressed genes than conventional methods. *Sci Rep* 12, 17438 (2022).
[3] Sanjiban Sekhar Roy, Y-h. Taguchi, Tensor Decomposition and Principal Component Analysis-Based Unsupervised Feature Extraction Outperforms State-of-the-Art Methods When Applied to Histone Modification Profiles,bioRxiv 2022.04.29.490081
[4] Y-H. Taguchi, Turki Turki, Principal component analysis- and tensor decomposition-based unsupervised feature extraction to select more reasonable differentially methylated cytosines: Optimization of standard deviation versus state-of-the-art methods, bioRxiv 2022.04.02.486807

2023/3/22〜25　オンライン