

拡張された非負値行列因子分解によるオミクスデータの解析

名古屋大学医学系研究科システム生物学分野, 阿部興
名古屋大学医学系研究科システム生物学分野, 島村徹平

1 はじめに

次世代シーケンス技術の発展により, 生体中に存在する様々な分子の情報を網羅的に取得することが可能になった. これらのデータの多くは高次元であるため, 実験結果の解釈のために次元圧縮の技術がよく用いられる. 中でも非負値行列因子分解 (NMF) および非負値テンソル因子分解は, メタゲノム解析 (Jiang *et al.*, 2012) や遺伝子発現解析 (Zhu *et al.*, 2017) をはじめとする多くのデータ分析への応用が報告されている. 本報告では, 複数のカテゴリカルな説明変数を自由にモデルに取り込むことができるよう拡張された非負値行列因子分解を提案する. 当日の報告では, 推定方法の詳細と応用例を示す.

2 モデル

あるサンプル n について非負の整数である応答変数のベクトル $\mathbf{y}_n = (y_{n,1}, \dots, y_{n,K})^\top$ とカテゴリカルな説明変数のベクトル $\mathbf{x}_n = (x_{n,1}, \dots, x_{n,K})^\top$ が観測されるとする. 形状パラメータ $a > 0$, レートパラメータ $b > 0$ のガンマ分布, および集中度パラメータ $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)^\top$ ($\alpha_k > 0$) のディリクレ分布を事前分布とした次のデータ生成過程を考える.

$$\begin{aligned}\lambda_{d,l} &\sim \text{Gamma}(a, b) \\ \mathbf{h}_l &\sim \text{Dirichlet}(\boldsymbol{\alpha}) \\ y_{n,k} &\sim \text{Poisson} \left(\tau_n \sum_{l=1}^L \left\{ \prod_{d=1}^D \lambda_{d,l}^{x_{n,d}} \right\} h_{l,k} \right),\end{aligned}$$

ここで $h_{l,k}$ はベクトル \mathbf{h}_l の第 k 成分を表し, L は分析者が設定する定数, τ_n はオフセット項である.

■例 3人の被験者の細菌の存在量を観測したとする. \mathbf{y}_n は被験者 n の細菌の存在量を表すベクトルである. このデータセット \mathcal{D} は次のように表す事ができる.

$$\mathcal{D} = \{(\text{被験者 1}, \mathbf{y}_1), (\text{被験者 2}, \mathbf{y}_2), (\text{被験者 3}, \mathbf{y}_3)\}$$

次のように被験者をダミー変数としてコード化すれば, 説明変数として扱う事ができる.

$$\mathbf{x}_1 = \{1, 0, 0\}, \mathbf{x}_2 = \{0, 1, 0\}, \mathbf{x}_3 = \{0, 0, 1\}$$

$\tau_n = 1$ とすれば, 提案モデルでは $y_{n,k}$ は次のように近似される.

$$y_{n,k} \approx \sum_{l=1}^L \left\{ \prod_{d=1}^3 \lambda_{d,l}^{x_{n,d}} \right\} h_{l,k} = \sum_{l=1}^L \lambda_{n,l} h_{l,k}.$$

これは NMF で行う分解と同等である. 提案モデルは NMF を特殊な場合として含む.

参考文献

- [1] Jiang, X., Weitz, J. S., & Dushoff, J. (2012). A non-negative matrix factorization framework for identifying modular patterns in metagenomic profile data. *Journal of Mathematical Biology*, 64(4), 697–711.
- [2] Zhu, X., Ching, T., Pan, X., Weissman, S. M., & Garmire, L. (2017). Detecting heterogeneity in single-cell RNA-Seq data by non-negative matrix factorization. *PeerJ*, 5, e2888.