

トピックモデルによる腸内細菌叢とパーキンソン病の関連分析

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

パーキンソン病は、腸内細菌叢との関連が示唆されている疾患の一つである. (e.g. Scheperjans *et al.*, 2015). しかし、分子メカニズムについては未知の部分が多く、詳細な解明には至っていない. 本研究では、次世代シーケンサーより得られたパーキンソン病患者のメタゲノムデータから、疾患と腸内細菌叢との関連性を分析するための統計モデルについて報告する. 本研究で扱うデータは、行方向をサンプル、列方向を各細菌叢とするデータ行列として表現できるが、概して 0 が多いスパース (疎) な行列になる. また、腸内細菌叢の構成はしばしば個人間で大きく異なり、いくつかのグループを形成することが知られている. このグループはエンテロタイプとも呼ばれる (Arumugam *et al.*, 2011). この 2 点の問題を解決するために、自然言語処理の分野で用いられるトピックモデルの応用により、データからエンテロタイプを推定すると同時に、パーキンソン病の影響を推定する手法を提案する. トピックモデルは文章の潜在的な「意味」を見出すために考案され、Woloszynek *et al.* (2017) では 16S-rRNA のデータへの適用も報告されている. 発表当日は、提案手法を 4 か国のパーキンソン病患者のメタゲノムデータに適用した結果について報告する.

2 モデル

y_{nk} を対象 n のバクテリア k の数とする. 観測されるのはこの y_{nk} を要素とする行列 $Y = (y_{nk})$ である. また Y の n 行目のベクトルを y_n と書くことにする.

以下の式 (1) で表される生成モデルを考える.

$$\begin{aligned} y_n | z_n, x_n, \alpha, \beta &\sim \text{Multinomial}(\text{softmax}(\gamma_{z_n} + \beta x_n)) \\ z_n | \pi &\sim \text{Categorical}(\pi) \\ \pi &\sim \text{Dirichlet}(a) \\ \alpha_k &\sim \text{Normal}(0, \rho^2) \\ \beta_k &\sim \text{Normal}(0, \sigma^2) \\ \gamma_l &\sim \text{Normal}_K(O_K, \tau I_K) \end{aligned} \quad (1)$$

z_n は対象 n の属するグループ (エンテロタイプ) であり未観測である. パラメータ π は混合比率, γ_l はエンテロタイプごとに異なる切片項. β_k はエンテロタイプにかかわらず共通のパーキンソン病の影響を示す. また O_L を L 次元の零行列, I_L は L 次元の単位行列とした.

これらのパラメータの推定にはラプラス近似を用いる. 本モデルを用いて分析を行うことで、パーキンソン病と関連する細菌を見つけ出すことができる.

参考文献

- [1] Arumugam, M. *et al.* (2011). Enterotypes of the human gut microbiome. *Nature*, 473 (7346), 174–180.
- [2] Scheperjans, F. *et al.* (2015). Gut microbiota are related to Parkinson's disease and clinical phenotype. *Movement Disorders*, 30(3), 350–358.
- [3] Woloszynek, S., Mell, J. C., Simpson, G., O'Connor, M. P., and Rosen, G. L. (2017). Uncovering thematic structure to link co-occurring taxa and predicted functional content in 16S rRNA marker gene surveys. *bioRxiv*, 146126.