

近年、細菌のゲノム情報を網羅的にシーケンスするメタゲノム解析によってヒト腸内細菌叢における各菌の組成割合に関するデータが得られるようになった。一方で、腸内細菌メタゲノムデータと表現型との関連を分析する手法は十分に確立されていない。その中で、自然言語処理において文書データを分析する手法であるトピックモデルを用いて細菌を分類する方法が提案されてきている。基本的なトピックモデルを拡張することにより表現型と関連する細菌集団を同定することが可能になると考えられる。本抄読会では、腸内細菌メタゲノムデータの基本的事項について述べ、トピックモデルの腸内細菌データへの応用方法について検討する。

#### 参考文献

1. DM Blei et al. (2003) Latent dirichlet allocation. *Journal of Machine Learning Research*, **3**,993–1022.
2. H Bisgin et al. (2011) Mining FDA drug labels using an unsupervised learning technique- topic modeling. *BMC Bioinformatics*, **12**, S11.