

審査の結果の要旨

氏名 奥井 佑

本研究は腸内細菌 16S rRNA データを用いて形質情報と関連する菌種を同定する関連解析において、いまだ標準化がなされていない、多変量データ解析手法を用いる際のデータの正規化法に着目し、rarefying 法に集団学習法を併用した手法を提案した。提案法と一般的な正規化法を用いた方法について、各種多変量データ解析手法を用いて性能をシミュレーション実験により比較評価したものであり、下記の結果を得ている。

1. 正規化法として組成割合と rarefying の性能を比較したところ、Random Forest(RF), Elastic Net(EN) では、組成割合を用いたモデルは rarefying 法を用いたモデルを菌種の同定力で上回り、菌種数が 100 の場合よりも 300 の場合において組成割合の rarefying に対する優位性は強くなった。また、EN について組成割合にバギングを併用した解析法についても、説明変数の数が 100 の場合にはバギングを併用しない解析法よりも同定力が低くなる現象がみられたが 300 の場合には相対的に同定力が高くなった。これにより、菌種数が多い階層のデータを用いる場合には組成割合を用いる際の変数間の擬似的相関の問題は低減することを示す結果となった。
2. EN と Sparse partial least squares discriminant analysis(SPLSDA)では提案法を用いることで他の解析法と比べて限定的ではあるが関連菌種の同定力が高くなり、特に菌種数が 100 の場合、各個体の総カウント数のばらつきが大きい場合に相対的な優位性が強くなった。よって、各個体の総カウント数のばらつきが大きい場合や菌種数が比較的少ない階層のデータを用いる場合には、説明変数間の関連を考慮した一般的な回帰モデルに提案法を併用することで形質・疾患の関連菌種をより正確に特定できる可能性を示唆する結果となった。
3. RF については組成割合を用いた解析法が最も関連菌種の同定力に優れており提案法の効果は限定的であるという結果となり、RF のモデルとしての特性に起因することが考えられた。EN と SPLSDA に提案法を適用した方法と比較して、RF を用いた解析法は同定力において低い値を示し、提案法を用いることで一般的な関連菌種の同定法の一つである RF の性能を上回ることができることを示唆する結果となった。

以上、本論文は腸内細菌 16S rRNA データを用いて形質情報と関連する菌種を同定する関連解析に関して各種正規化法について性能評価を行うとともに、今後より詳細な検証が必要ではあるものの提案法が一般的に用いられている解析法の性能を上回ることをシミュレーション実験により示した基礎的な研究であることから腸内細菌 16S rRNA データを用いた関連解析の研究に貢献をなすと考えられ、学位の授与に値するものと考えられる。