DeepBiome: a phylogenetic tree regularized deep neural network for microbiome data analysis

Jin Zhou

University of Arizona
E-mail: jzhou@email.arizona.edu

Abstract: Evidence linking microbiome and human health is rapidly growing, suggesting that the microbiome profile may serve as a novel predictive biomarker for diseases. Two fundamental challenges of analyzing microbiome data are that bacteria count tables are very sparse and bacteria are classified at a hierarchy of taxonomic levels, ranging from species to phylum. Existing statistical and computational tools often focus on identifying the microbiome association either at the community level or at a specific pre-defined taxonomic level to aggregate rare features. They fail to incorporate hierarchical structure information and cannot learn from the data to aggregate microbiome contribution, therefore, leading to inaccurate selection and prediction.

We present DeepBiome, a deep learning model, to uncover the microbiome-phenotype association network and visualize its path to disease. The proposed DeepBiome takes microbiome abundance data as input and uses the phylogenetic taxonomy to guide the decision of the optimal number of layers and neurons in the deep learning architecture. By doing so, we relieve the computation burden of tuning hyperparameters often encountered in the general deep learning architecture. In addition, we introduce a phylogeny regularized weight decay technique to regularize DeepBiome model and avoid overfitting. DeepBiome analyzes the whole microbiome profile and its path to disease and identifies taxa associated with outcome at each taxonomic level. Our algorithm is applicable to both regression and classification problems in microbiome data analysis. The simulation studies and real data analysis show that DeepBiome is a highly accurate, efficient method and provides deep understanding of complex microbiome-phenotype association.