

Ensemble methods for classification of patients for personalized medicine with high-dimensional data

Hojin Moon

Division of Biometry & Risk Assessment
U.S. Food and Drug Administration (FDA)

Abstract

Classification methods are commonly used for prediction of response to therapy to help individualize clinical assignment of treatment. The methods are required to be highly accurate for optimal treatment on each patient. Typically, there are numerous genomic and clinical variables over a relatively small number of patients, which presents challenges for most traditional classification methods to avoid over-fitting the data. We developed a robust classification method for high-dimensional data based on ensembles of classifiers from the optimal number of random partitions. The proposed method is applied to genomic data sets on lymphoma patients and lung cancer patients to distinguish disease subtypes for optimal treatment and to genomic data on breast cancer patients to identify patients most likely to benefit from adjuvant chemotherapy after surgery. The performance of the proposed method is consistently good or better compared to the other classification algorithms. We find that the predictive accuracy can be improved by adding some relevant demographic, clinical and/or histopathological measurements to the genomic data.