Subgroup Discovery Using Consensus Clustering Methods

J. Richard Landis

University of Pennsylvania
E-mail: jrlandis@pennmedicine.upenn.edu

Abstract: Discovery of patient subtypes with differential profiles of risk for multiple outcomes is essential for improving health promotion and precision medicine strategies. A K-means clustering algorithm, using an average linkage method, was applied to I item clusters as variables, and was repeated within 1,000 randomly selected subsamples of size 80%N, utilizing R Bioconductor ConsensusClusterPlus (Wilkerson, Hayes and Neil (2010)). An evaluation criterion based on the "Proportion of Ambiguous Pairs within Clusters" (PAC) was used to generate a mean consensus score for each of the K clusters. A new sequential hypothesis testing approach was utilized to determine the ideal number of clusters. These methods were applied to deep phenotyping data from the Multidisciplinary Approach to the Study of Chronic Pelvic Pain (MAPP) Research Network (http://www.mappnetwork.org/).