
JOSHUA ARMSTRONG, Dalhousie University

Application of three data mining and machine learning methods in data from the Honolulu-Asian Aging Study

As the population of older adults continues to expand across the globe, health care systems will face patient populations that are increasingly complex and difficult to manage. The challenges involved in caring for older individuals arise due to an increased number of health deficits that commonly accumulate with age. In order to better understand the accumulation of health deficits in aging populations, data mining and machine learning methods can be applied to detect previously unidentified patterns in clinical characteristics collected in longitudinal studies of aging. In a secondary analysis of data from the Honolulu-Asian Aging Study, accumulation of deficits frailty index (FI) scores were calculated for 3651 men, aged 71-93. This presentation will examine these accumulated deficits and illustrate three methods: (1) K-means clustering, (2) association rule mining (apriori algorithm), and (3) Random forests. Cluster analyses focused on identifying common patterns of deficits for the frailest individuals (FI \geq 0.35). Association rule mining, also known as market basket analysis, was applied to identify the most common combinations of deficits found in this sample of older men. Using data collected on these men 30 years previously, mid-life characteristics associated with fitness (low level of frailty: FI \leq 0.10) were identified and ranked using the Random forest algorithm. While novel exploratory methods should not replace the traditional statistical techniques that are frequently used in health research, data mining and machine learning methods should be developed for use in health data as they may be able to provide insights into the complex problems of gerontology and geriatrics.