Abstract

Cluster Analysis on Longitudinal Cognitive Function Trajectories:
A Poisson Mixture Model Approach

Qianqian Zhu
Master of Science
Graduate Department of Biostatistics Division at Dalla Lana School of Public Health
University of Toronto
2018

Dementia has been one of the fastest growing public health issues for the past several decades because of the aging of the world’s population. It not only does make people suffer, but also becomes a serious burden on caregivers. Cognitive function decline is a normal phenomena of aging as well as a symptom of dementia. Analyzing how individuals cognitive function change can help scientists understand and possibly prevent the progression of dementia. Cognitive Performance Scale (CPS), which is derived from the Minimum Data Set (MDS), is a seven-category scale that indicates the cognitive impairment of a person. Motivated by other researchers, we aim to find distinct CPS change trajectories among individuals who reside in long term care facilities in the province of Ontario.

Cluster analysis aims to group data into subsets that contain similar objects. One of the well-known issues of cluster analysis is the validation of the clustering results. Therefore, we also intend to simulate a dataset which has similar characteristics to the real one in order to validate the results obtained via some clustering algorithms. Thus, we are also interested in estimating the transition probabilities, initial proportions and template proportions.

Finite mixture model is a model-based clustering method to discover latent subpopulations among an overall population. Poisson mixture model, a special case of the finite mixture models, was introduced into our study to help us find the heterogeneous subgroups since we are provided with the aggregate count data across individuals. It can also help us find the probabilities and proportions that we target to estimate. We designed a method to define the longitudinal trajectories by the transition probability matrix. Furthermore, we parameterized the functional form of the log-linear model for each component in such a way that the relationship between the
regression coefficients and the probabilities that we aim to acquire makes sense and is simple to transform. In addition, we proposed the restricted Poisson mixture model, which modifies the conventional EM algorithm of Poisson mixture models by adding some constraints to the parameters in order to obtain the distinct templates.

Applications of both the k-means algorithm and a longitudinal clustering method (Leffondré et al., 2004) were implemented on the data simulated with the probabilities and proportions estimated from the count data. Results suggested that the longitudinal clustering method is more appropriate as it not only has a higher accuracy rate but also has the ability to capture the trajectory change patterns.