CLUSTERING FOR SPARSELY SAMPLED LONGITUDINAL DATA BASED ON BASIS EXPANSIONS

Michio Yamamoto[1] and Yoshikazu Terada[2]

[1] Graduate School of Human Sciences, Osaka University / RIKEN AIP, (e-mail: yamamoto.michio.hus@osaka-u.ac.jp)

[2] Graduate School of Engineering Science, Osaka University / RIKEN AIP, (e-mail: terada.yoshikazu.es@osaka-u.ac.jp)

ABSTRACT: In longitudinal data, the observations often occur at different time points for each subject. In such a case, ordinary clustering algorithms like K-means clustering cannot be applied directly. Instead, one may apply a smoothing technique to get individual continuous trajectories, followed by finding groups among the trajectories using some clustering algorithm. However, this is inappropriate when each subject's data are observed at only a few time points. Thus, we develop a new clustering algorithm for sparsely sampled longitudinal data, which can be considered a natural extension of the K-means clustering. We show the consistency of the proposed estimator under mild regularity conditions. We also evaluate its performance through simulation studies and data applications.

KEYWORDS: clustering, longitudinal data, functional data analysis