

Non-parametric clustering of longitudinal functional data with application to NMR spectra of 18 kidney transplant patients.

Minzhen Xie¹, Jeanine Houwing-Duistermaat², Haiyan Liu³

¹ Department of Statistics, University of Leeds, Leeds, UK

² Department of Statistics, University of Leeds, Leeds, UK

³ Department of Statistics, University of Leeds, Leeds, UK

E-mail for correspondence: mmmxi@leeds.ac.uk

Abstract: The amount of data in the health domain is growing rapidly, which have different forms: omics, imaging and functional. A relevant question is whether these data can be used to cluster subjects to reveal their underlying health status. The aim of this research is to cluster the 18 kidney transplant patients based on their NMR spectra. An NMR spectrum is a function in mass, and the NMR spectra of each patient are recorded up to nine times (longitudinal design). It might loss relevant information, if we apply multivariate clustering method such as K-means on extracted scalar summaries of longitudinal measured NMR spectra.

To use all available information, we propose a non-parametric clustering method for multivariate functional data. Firstly, the distance of multivariate curves is defined and small ball probability is computed based on defined distance. Secondly, we compute the heterogeneity of original sample and partitioned samples by using mean, median and mode. Thirdly, we define a criterion to determine whether the obtained clustering provides homogeneous clusters or require further splits. We compare our method with other (non) functional clustering methods via simulation and apply the method to the kidney data.

Key words: Nonparametric clustering; Longitudinal functional data

Phikippe Vieu, Frédéric Ferraty (2006). Nonparametric Functional Data Analysis. New York: Springer-Verlag.

Ana-Maria Staicu, So Young Parka (2012). Longitudinal Functional Data Analysis. The ISI's Journal for the Rapid 4.