Rivera-García, Diego; García-Escudero, Luis A.; Mayo-Iscar, Agustín; Ortega, Joaquín

Robust clustering for functional data based on trimming and constraints. (English)

Zbl 1474.62166

Summary: Many clustering algorithms when the data are curves or functions have been recently proposed. However, the presence of contamination in the sample of curves can influence the performance of most of them. In this work we propose a robust, model-based clustering method that relies on an approximation to the “density function” for functional data. The robustness follows from the joint application of data-driven trimming, for reducing the effect of contaminated observations, and constraints on the variances, for avoiding spurious clusters in the solution. The algorithm is designed to perform clustering and outlier detection simultaneously by maximizing a trimmed “pseudo” likelihood. The proposed method has been evaluated and compared with other existing methods through a simulation study. Better performance for the proposed methodology is shown when a fraction of contaminating curves is added to a non-contaminated sample. Finally, an application to a real data set that has been previously considered in the literature is given.

MSC:
62G35 Nonparametric robustness
62H25 Factor analysis and principal components; correspondence analysis
62H30 Classification and discrimination; cluster analysis (statistical aspects)
62R10 Functional data analysis
68T10 Pattern recognition, speech recognition

Keywords:
functional data analysis; clustering; robustness; trimming; functional principal components analysis

Software:
fdau.sc; TCLUST; funHDDC; fda (R); Funclustering

Full Text: DOI arXiv

References:


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.