A CLUSTERING METHOD FOR DISTRIBUTIONAL DATA BASED ON A LDQ TRANSFORMATION

Rosanna Verde [1], Gianmarco Borrata [2] and Antonio Balzanella [1]

ABSTRACT: This work deals with a clustering method for distributional data. The set of objects to be clustered are described by p distributional variables. Each object is represented by p density probability functions (*dpf*'s), or empirical ones. In consideration of the most recent developments in distributional data analysis (DDA), we introduce a transformation of the quantile functions, *qf*'s, associated to the *dpf*'s, in Logarithm Derivative Quantiles (LDQ) functions, which allows to map density probability functions in an Hilbert space. Our proposal is based on a Dynamic Clustering Clustering type-algorithm, where the centroid of the clusters are represented by linear combination of LDQ functions; the objects are assigned to the clusters according to minimum sum of the squared distance from the centroid function. Applications on synthetic and real data have corroborated the new method.

KEYWORDS: symbolic data analysis, distributional data, quantile density functions

[1] Department of Mathematics and Physics, University of Campania Luigi Vanvitelli, (e-mail: `rosanna.verde@unicampania.it`, `antonio.balzanella@unicampania.it`)

[2] Department of Social Science, University of Naples Federico II, (e-mail: `gianmarco.borrata@studenti.unicampania.it`)