On the performance of K-means clustering for dissimilarity data. Application in the dynamic time warping framework for time series

A seminar by José Fernando Vera Vera

Department of Statistics and Operational Research
Universidad de Granada (Spain)

Friday 19 Apr 2024 | 12.30 p.m.
Room Sc 60
Department of Statistical Sciences

Partitioning algorithms, and in particular K-means clustering, are widely used in time series analysis. K-means clustering is intrinsically related to the use of the Euclidean distance as a measure of dissimilarity. When other dissimilarity measures are involved, K-means clustering is usually replaced by the optimisation of a sums-of-the-stars clustering criterion, giving rise to an algorithm other than that of K-means, such as K-medoids. Another common restriction in the implementation of K-means concerns the need to estimate the average as the cluster prototype, which may represent a drawback for this method in some applications as for time series clustering, when elastic measures such as dynamic time warping are used. We propose a multidimensional scaling based K-means clustering algorithm that enables the use of K-means clustering together with any dissimilarity measure. In particular, the proposed procedure can be used with dynamic time warping, without requiring us to estimate cluster prototypes for the time series. This procedure is a true K-means clustering algorithm that searches for the partition in an equivalent auxiliary configuration, usually in a dimension lower than the time series length. The approach proposed is of particular interest when dynamic time warping is used in the analysis of series of unequal length and/or when some data are missing, and hence Euclidean distances cannot be used.

Università degli Studi di Padova

Dipartimento di Scienze Statistiche