Assessment of a partition and its clusters according to a Rand index-based stability criterion

Bertrand Patrice*
Université Paris-Dauphine, Ceremade, Paris, France - bertrand@ceremade.dauphine.fr

El Moubarki Lassad
Université de Sfax, Sfax, Tunisia - elmoubarki.lassad@yahoo.fr

Bel Mufti Ghazi
ESSEC-Tunis, Larime, Tunis, Tunisia - belmufti@yahoo.com

Measures of stability are currently used in order to assess the results of partitioning methods. Frequently proposed measures of stability are estimated central tendencies of the distribution of the dissemblance degree between two partitions provided by the same partitioning procedure applied to two randomly perturbed versions of the original dataset. A simple and efficient way to compare two partitions being provided by the adjusted Rand index, this index is often employed to assess partitioning stability. In this talk, we first introduce a partitioning stability measure which is derived directly from the Rand index, and then propose two barycentric decompositions of this Rand index-based measure of stability. The obtained components of the two decompositions are interpreted as measures of cohesion, isolation and validity of each cluster, and by correcting for chance these measures, we introduce the corresponding adjusted indexes that evaluate these characteristics for each individual cluster. These results leads us to propose several partitioning validity indexes which aim to determine the optimal number of clusters of a partition, some of them being based on the bi-criterion of both isolation and cohesion of each individual cluster. Experimentations achieved on simulated and real data sets show that these indexes of cluster validity compare favorably with most successful methods for predicting the number of clusters.

Keywords: Partitioning stability; Cluster validity; Resampling; Adjusted Rand index.