Mach Learn
DOI 10.1007/s10994-013-5394-z

Unsupervised ensemble minority clustering

Edgar Gonzàlez · Jordi Turmo

Received: 16 May 2012 / Accepted: 8 June 2013

Abstract Cluster analysis lies at the core of most unsupervised learning tasks. However, the majority of clustering algorithms depend on the all-in assumption, in which all objects belong to some cluster, and perform poorly on minority clustering tasks, in which a small fraction of signal data stands against a majority of noise.

The approaches proposed so far for minority clustering are supervised: they require the number and distribution of the foreground and background clusters. In supervised learning and all-in clustering, combination methods have been successfully applied to obtain distribution-free learners, even from the output of weak individual algorithms.

In this work, we propose a novel ensemble minority clustering algorithm, EWOCS, suitable for weak clustering combination. Its properties have been theoretically proved under a loose set of constraints. We also propose a number of weak clustering algorithms, and an unsupervised procedure to determine the scaling parameters for Gaussian kernels used within the task.

We have implemented a number of approaches built from the proposed components, and evaluated them on a collection of datasets. The results show how approaches based on EWOCS are competitive with respect to—and even outperform—other minority clustering approaches in the state of the art.

Keywords Clustering · Minority clustering · Ensemble clustering · Weak learning

Editors: Emmanuel Müller, Ira Assent, Stephan Günneman, Thomas Seidl, Jennifer Dy.

E. Gonzàlez (✉) · J. Turmo
TALP Research Center, Universitat Politècnica de Catalunya, c/Jordi Girona, 1, 08034 Barcelona, Spain
e-mail: egonzalez@lsi.upc.edu

J. Turmo
e-mail: turmo@lsi.upc.edu

Present address:
E. Gonzàlez
Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA

Published online: 03 July 2013