ID de Contribution: **16**   Type: **Non spécifié**

MASSICCC: A SaaS Platform for Clustering Mixed Data

vendredi 22 juin 2018 14:00 (45 minutes)

The "Big Data" paradigm involves large and complex data sets where the clustering task plays a central role for data exploration. For this purpose, model-based clustering has demonstrated many theoretical and practical successes in a various number of fields. In this context, user-friendly software are essential for speeding up diffusion of such academic advance inside the applicative world. MASSICCC (massive clustering in cloud computing) is a user-friendly SaaS platform which hosts three software specialized in different clustering tasks and written in C++. This platform allows to manipulate complex data with very light computing tools (as a smartphone), including also some dynamical graphical outputs. However, it offers also the possibility to export the results into a R data format for further more expert tasks. The three embedded software are Mixmod, Mixtcomp and Blockcluster. Mixmod (Lebret et al. 2015) is dedicated to clustering of continuous, categorical and a mixing of continuous and categorical data. Mixtcomp (Biernacki 2015) adds the possibility to cluster totally mixed data (continuous, categorical, count, ordinal, rank, functional), potentially including missing or partially missing (like interval) data. Blockcluster (Bhatia et al. 2017) is dedicated to co-clustering of large data sets composed of different kinds of data like continuous, categorical and count ones. In this talk, we will make a focus on both the Mixmod and MixtComp software.
MASSICCC is freely available at https://massiccc.lille.inria.fr

References:
P. Bhatia, S. Iovleff & G. Govaert (2017). Blockcluster: An R Package for Model-Based Co-Clustering. Journal of Statistical Software, 76:9.
C. Biernacki (2015). Model-based clustering with mixed/missing data using the new software MixtComp. 8th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2015), University of London, UK, 12-14 December.
R. Lebret, S. Iovleff, F. Langrognet, C. Biernacki, G. Celeux & G. Govaert (2015). Rmixmod: The R Package of the Model-Based Unsupervised, Supervised and Semi-Supervised Classification Mixmod Library. Journal of Statistical Software, 67:6.

Auteur principal: Prof. BIERNACKI, Christophe (Université Lille 1/INRIA)

Orateur: Prof. BIERNACKI, Christophe (Université Lille 1/INRIA)

Classification de Session: Softwares