Latest developments of the SYR software for Symbolic Data Analysis of complex and big data

Filipe Afonso1,*, Raja Haddad1,2, Edwin Diday2

1. SYROKKO Company, Aéropôle - Bat. Aéronef, 95731 Roissy CDG Cedex, France
2. Paris Dauphine University, 75775 Paris Cedex 16, France
*Contact author: afonso@syrokko.com

Keywords: Symbolic data analysis software, Big Data, Complex data, Clustering and Visualizing symbolic data

We present the main evolutions of the SYR software for Symbolic Data Analysis of Complex Data (Afonso et al., 2014). The SYR software is a SYROKKO company product. Its aim is to extract, from a data file, up to several millions of units, a reduced number of units called “concepts” which summarize the initial data. These units are described by standard categorical or numerical variables, as well as by interval variables, multi-valued variables and by bar-chart and histogram-valued variables. These new kinds of variables allow keeping the internal variation of each concept.

The presentation focus on the recent developments extended to any kind of symbolic data (Decision trees, Principal Component Analysis) and on ongoing research and development concerning the extension of the software to Big Data (Du Mouza et al., 2009). The software is presented through industrial applications (for example, Nuemi et al., 2013, Diday et al. 2013) and comparisons with the academic SODAS software for symbolic data (Diday and Noirhomme, 2008) are given.

References


