New Advances in Symbolic Data Analysis and Spatial Classification

Edwin Diday

Ceremade, Université Paris-Dauphine
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16, France
diday@ceremade.dauphine.fr

Abstract The usual Data mining model is based on two parts: the first concerns the units (called here *individuals*), the second, contains their description by several standard variables including a class variable. The Symbolic Data Analysis model needs two more parts: the first concerns units called "concepts" and the second concerns their *description*. The concepts are characterized by a set of properties called "intent" and by an "extent" defined by the set of individuals which satisfy these properties. These concepts are described by *symbolic data* which are standard categorical or numerical data and moreover interval, histograms, sequences of values, etc. These new kind of data allows keeping the internal variation of the extent of each concept. Then, new knowledge can be extracted from this model by new tools of Data Mining extended to concepts considered as new units. Among these tools, Spatial Classification allows a graphical visualisation of the given concepts on a grid and at different level of generalisation organised by a spatial hierarchy or pyramid (allowing overlapping clusters). The SYR software has been developed by SYROKKO company after the academic SODAS software developed by two EUROPEAN projects until 2003, The first aim of SYR is to extract, from a data file (.txt, .csv, ACCESS database) of several millions of units a reduced number of units which are *concepts* summarizing the initial data. Then SYR can create handle (select, cut, move rows or columns.) and visualise a symbolic data file thanks to user-friendly graphical output. Finally SYR produces new knowledge by Symbolic Data Analysis tools.

References

[1] Billard L. and Diday E., Symbolic Data Analysis: conceptual statistics and data Mining. Wiley. ISBN 0-470-09016-2. 351 pages (2006).

[2] Diday E. and Noirhomme M., Symbolic Data Analysis and the SODAS software. Wiley. ISBN 978-0-470-01883-5, 457 pages (2008) .

[3] Diday E., Spatial classification. DAM (Discrete Applied Mathematics) Volume 156, Issue (2008).