Alpha Galois Lattices for conceptual clustering

Nathalie Pernelle[1], Henry Soldano[2,3], and Véronique Ventos[1]

[1] Université Paris Sud - CNRS(LRI) & INRIA(Futurs)
LRI, Bat. 490,
91405 Orsay, France
{pernelle, ventos}@lri.fr
[2] Université Paris-Nord
LIPN, UMR-CNRS 7030, av. J.B. Clément
93430 Villetaneuse, France
soldano@lipn.univ-paris13.fr
[3] Université Paris VI
Atelier de Bio-Informatique, Dpt des Sciences de la Vie, 12 rue Cuvier
75005 Paris, France

Abstract

Our work deals with the representation of data by clustering them in a small number of classes organized in a hierarchy and described at an appropriate level of abstraction. Our basic representation is a Galois lattice, i.e a lattice in which terms of a representational language are partitioned into equivalence classes w.r.t their *extent*. The main drawback of this structure is that the size of such a lattice can be very large when dealing with real-world data sets. In our approach, the basic ideas are the use of different languages of description of classes (i.e. languages more or less expressive to describe clusters of instances) and the use of basic classes of instances (sets of instances sharing the same type) in order to modify the definition of the *extent*. In this paper, we only focus on this modification: an instance now belongs to the $\alpha - extent$ of a term c when it belongs to the *extent* of c and when at least α % of the instances sharing the same type also belongs to the extent of c. This allows to deal either with some representational noise in instances (with high values of α) or with heterogeneous user-defined basic classes (using small value of α). When $\alpha = 100$, a term represents a cluster of basic classes and when $\alpha = 0$, the $\alpha - extent$ corresponds to the usual notion of *extent*. An $\alpha - extent$ induces an alpha Galois Lattice which is coarser than the usual Galois Lattice since the set of nodes of an alpha Galois Lattice is a subset of the set of nodes of the usual Galois Lattice.