Alpha Galois Lattices: An Overview

Véronique Ventos[1] and Henry Soldano[2]

[1] LRI, UMR-CNRS 8623, Université Paris-Sud, 91405 Orsay, France
ventos@lri.fr
[2] L.I.P.N, UMR-CNRS 7030, Université Paris-Nord,
93430 Villetaneuse, France
soldano@lipn.univ-paris13.fr

Abstract. What we propose here is to reduce the size of Galois lattices still conserving their formal structure and exhaustivity. For that purpose we use a preliminary partition of the instance set, representing the association of a "type" to each instance. By redefining the notion of *extent* of a term in order to cope, to a certain degree (denoted as α), with this partition, we define a particular family of Galois lattices denoted as *Alpha Galois lattices*. We also discuss the related implication rules defined as inclusion of such α-extents and show that Iceberg concept lattices are Alpha Galois lattices where the partition is reduced to one single class.

1 Introduction

Galois lattices (or concept lattices) are well-defined and exhaustive representations of the concepts embedded in a data set since they allow us to obtain every subset of instances distinguishable according to the chosen attributes. However, when dealing with real-world data sets the size of such a lattice can be too large to be handled. Various techniques have been proposed to reduce the size of concept lattices by eliminating part of the nodes (e.g. [7]). In particular, Iceberg concept lattices [14, 17] represent the topmost part of a concept lattice w.r.t. a global criterion of frequency: only nodes with an *extent* cardinality satisfying a threshold according to the whole data set are kept. In this paper, we present more flexible Galois lattices in which the number of nodes is controlled according to a local criterion of frequency linked to a prior partition of the set of instances.

The partition is a set of *basic classes* which are clusters of instances sharing the same basic type. For instance, in real data concerning the electronic catalog of computer products C/Net (http://www.cnet.com), there are 59 different basic types (e.g. *Laptops*, *HardDrives*, *NetworkStorage*) for 2274 instances. Basic classes are then used in order to add a local criterion of frequency to the notion of *extent* as follows: an instance i now belongs to $ext_\alpha(T)$, the α-*extent* of a subset T of the set of attributes, when it belongs to $ext(T)$, the extent of T, (i.e. i has every of T's properties), and when at least α % of the instances of the basic class of i also belong to $ext(T)$. This new notion of α-*extent* is used in the

B. Ganter and R. Godin (Eds.): ICFCA 2005, LNAI 3403, pp. 299–314, 2005.

Galois connection related to the family of *Alpha Galois lattices*. Alpha Galois lattices were first introduced in [12] as a part of the system ZooM.

In comparison with concept lattices, Alpha Galois lattices are mainly characterized by the following properties:

- For the same set of attributes, the same set of individuals, and for any value of α, the Alpha Galois lattice G_α is coarser than the concept lattice G, i.e. the set of nodes of G_α is a subset of the set of nodes of the concept lattice G.
- G_0 exactly is G, and G_{100} also is a concept lattice built from a set of instances that each represents one basic class.
- The values of α define a total order on *Alpha Galois lattices* where the *Alpha Galois lattice* induced by $ext_{\alpha 1}$ is coarser than the *Alpha Galois lattice* induced by $ext_{\alpha 2}$ if $\alpha 1 \geq \alpha 2$.
- When all individuals belong to a single basic class, the corresponding Alpha Galois lattice is an Iceberg concept lattice where $\frac{\alpha}{100} = minsupp$.
- A property (i.e. an attribute) can belong to an *intent* of an Alpha Galois lattice G_α even if it is not globally frequent. For instance, in G_{90} the "support" property will appear since in the $HardDrives$ basic class, 92 % of the instances of $HardDrives$ were sold with support. Actually, this property is not globally frequent (13 products out of 2274, i.e. 0.5 %) and so would not apppear in the corresponding Iceberg concept lattice with $minsupp = 0.9$
- The inclusion of α-*extent* corresponds to particular implication rules, representing some kind of approximation of usual implication rules, that depends on the selected partition of the instances.

The general framework of Galois lattices is given in section 2. In section 3, we present Alpha Galois lattices illlustrated with a simple example. Section 4 presents experimental results on the C/net data set and discusses the ability of such a representation to deal with exceptional data (α near 0 or near 100). Section 5 first discusses Iceberg Alpha Galois lattices together with α-implication rules, and then briefly addresses theoretical issues as the nature of the objects of a formal context which concept lattice is isomorphic to an Alpha Galois lattice. Finally, related work and future work are discussed in section 6.

2 Preliminaries and Definitions

Detailed definitions, results and proofs regarding Galois connections and lattices may be found in [1, 2]. Other results concerning Galois lattices in the field of Formal Concept Analysis can be found in [4]. However we need a more general presentation than the one in [4] as our main goal is to construct Galois lattices where the notion of *extent* is not the usual one. In the rest of the paper we denote as Galois lattice the formal structure that we define hereunder and we will denote as concept lattice the Galois lattice as presented in [4].We consider in our presentation that the reader is familiar with the definitions of *ordered set* and *lattice*. We also recall that a mapping w from an ordered set M to M