A local discretization of continuous data for lattices: *Technical aspects*

Nathalie Girard, Karell Bertet and Muriel Visani

Laboratory L3i - University of La Rochelle - FRANCE
ngirar02, kbertet, mvisani@univ-lr.fr

Abstract. Since few years, Galois lattices (GLs) are used in data mining and defining a GL from complex data (*i.e.* non binary) is a recent challenge [1,2]. Indeed GL is classically defined from a binary table (called context), and therefore in the presence of continuous data a *discretization* step is generally needed to convert continuous data into discrete data. Discretization is classically performed before the GL construction in a **global** way. However, **local** discretization is reported to give better classification rates than global discretization when used jointly with other symbolic classification methods such as decision trees (DTs). Using a result of lattice theory bringing together set of objects and specific nodes of the lattice, we identify subsets of data to perform a **local discretization for GLs**. Experiments are performed to assess the efficiency and the effectiveness of the proposed algorithm compared to global discretization.

1 Discretization process

The discretization process consists in converting continuous attributes into discrete attributes [3]. This conversion can induce scaling attributes or **disjoint intervals**. We focus on the latter. Such a transformation is necessary for some classification models like symbolic models, which cannot handle continuous attributes [4]. Consider a continuous data set $D = (O, F)$, where each object in O is described by p continuous attributes in F. The discretization process is performed by iteration of attribute splitting step, according to a **splitting criterion** (Entropy [3], Gini [5], χ^2 [6], ...) until a **stopping criterion** S is satisfied (a maximal number of intervals to create, a purity measure,...).
More formally for one discretization step, for selecting the best attribute to be cut, let $(v_1, \ldots, v_N)$ be the sorted values of a continuous attribute $V \in F$. Each v_i corresponds to a value verified by one object of the data set D. The set of possible cut-points is $C_V = (c_V^1, \ldots, c_V^{N-1})$ where $c_V^i = \frac{v_i + v_{i+1}}{2}$ $\forall i \leq N-1$.
The best cut-point, denoted c_V^*, is defined by:

$$c_V^* = argmax_{c_V^i \in C_V}(gain(V, c_V^i, D)) \quad (1)$$

where $gain(V, c, D)$ denotes in a **generic** manner the **splitting criterion** computed for the attribute V, the cut-point $c \in C_V$ and the data set D.
The best attribute, denoted V^*, is the $V \in F$ maximizing the **splitting criterion** computed for its best cut-point (*i.e.* c_V^*):

$$V^*(D) = argmax_{V \in F}(gain(V, c_V^*, D)) \quad (2)$$

Finally for one discretization step, the attribute V^* is divided into two intervals: $[v_1, c^*_{V^*}]$ and $]c^*_{V^*}, v_n]$ and the process is repeated.
This process can be run using, at each step, all the objects in the training set. This is **global discretization**. It can also be run during model construction considering, at each step, only a part of the training set. This is **local discretization**. In [7], Quinlan shows that **local discretization improves supervised classification** with decision trees (DTs) as compared with global discretization. In DT construction, the growing process is iterated until S is satisfied. Local discretization is performed on the subset of objects in the current node to select its best attribute ($V^*(node)$), according to the splitting criterion. Given the structural links between DTs and Galois lattices (GLs) [8], we propose a local discretization algorithm for GL and compare its performances with a global discretization.

2 Local discretization for Galois lattices

A GL is generally defined from a binary relation R between objects O and binary attributes I - *i.e.* a binary data set also called a **formal context** - denoted as a triplet $T = (O, I, R)$. A GL is composed of a set of **concepts** - a concept (A, B) is a maximal objects-attributes subset in relation - ordered by a generalization/specialization relation. For more details on GL theory, notation and their use in classification tasks, please refer to [9,10]. To define a local discretization for GL, we have to identify at each discretization step the subset of concepts to be processed. Given a subset of objects $A \in P(O)$, there always exists a smallest concept M containing this subset and identified in lattice theory as a **meet-irreducible concept** of the GL [11]. Moreover, it is possible to compute the set of meet-irreducibles directly from the context, thus the generation of the lattice is useless [12]. Consequently, local discretization is performed on the set of meet-irreducible concepts MI which does not satisfy S. Attributes in MI are locally discretized: the best attribute $V^*(M)$ for each $M \in MI$ is computed according to eq. (3); then the best one $V^*(MI)$ (eq. (4),(5)) for the whole set MI is split into two intervals as explain before. The context T is then updated with these new intervals; and its MI are computed. The process is iterated until all $M \in MI$ verify the stopping criterion S. The context T is initialized with, for each continuous attribute, an interval -*i.e.* a binary attribute- containing all continuous values observed in D; thus each object is in relation with every binary attributes of T. The GL of the inital context T contains only one concept (O, I) being a meet-irreducible concept, which is used to initialize MI. See [13] for more details on the algorithm.
The main difference with DT is that splitting an attribute in a GL impacts all the other concepts of the GL that contain this attribute, and due to the order relation between concepts $\leq$, the structure of the GL is also modified. Whereas, when an attribute is split in a DT node, predecessors and others branches are not impacted. In order to select the best $V^*(MI)$ over all the concepts sharing this attribute, we introduce different computing of $V^*(MI)$.

Let $MI = \{D_q = (A_q, B_q); q \leq Q\}\}$ be the set of meet-irreducible concepts not satisfying S. The best attribute $V^*(D_q)$ associated to its best cut-point is first computed for each concept $D_q \in MI$:

$$V^*(D_q) = argmax_{V \in B_q}(gain(V, c_V^*, D_q)) \quad (3)$$

where c_V^* is defined by (1) for D_q instead of D.
Let us define $I_{MI}^* = \{V^*(D_1), \ldots, V^*(D_Q)\}$ the set of best attributes associated to each concept in MI. The best attribute $V^*(MI)$ among I_{MI}^* can be defined in two different ways:
By local discretization: Local discretization selects the best attribute $V \in I_{MI}^*$ as the one having the best gain for MI:

$$V^*(MI) = argmax_{V^*(D_q) \in I_{MI}^*}(gain(V^*(D_q), c_{V^*(D_q)}^*, D_q)) \quad (4)$$

By linear local discretization: Linear local discretization takes into account that the split of one attribute $V \in I_{MI}^*$ in a concept D_q can impact the other concepts. So we compute a linear combination of the criterion as the sum of the gain for each concept $D_{q'} \in MI$ containing this attribute V. The selected attribute is the one that gives the best linear combination:

$$V^*(MI) = argmax_{V \in I_{MI}^*}(\sum_{D_{q'} \in MI | V \in B_{q'}} \frac{|A_{q'}|}{\sum_{D_q \in MI} |A_q|} * gain(V, c_V^*, D_{q'})) \quad (5)$$

3 Experimental comparison

The study is performed on three supervised databases of the UCI Machine Learning Repository[1]: the Image Segmentation database (Image1), the Glass Identification Database (GLASS) and the Breast Cancer Database (BREAST Cancer). We also use one supervised data set stemming from GREC 2003 database[2] described by the statistical Radon signature (GREC Radon). Table 1 presents the **complexity of each lattice structure** associated to each discretization algorithm and the **classification performance** using each GL by navigation [14] and using CHAID as DT classifier [6]. Discretization is performed in each case with χ^2 as a splitting and stopping supervised criterion.

4 Conclusion

The study [3] shows that for DTs, local discretization induces more complex structures compared to global discretization; Table 1 shows that **for GL, on the contrary, local discretization allows to reduce the structures' complexity**. In [7], Quinlan proves that local discretization improves classification performance of DTs compared to global discretization; as in DTs, Table 1 shows that **local discretization improves GLs classification performances**.

[1] http://archive:ics:uci:edu/ml [2] www.cvc.uab.es/grec2003/symreccontest/index.htm

Table 1. Structures complexity and Classification performance

	Nb concepts			Recognition rates			
	Local	Linear Local	Global	Local	Linear Local	Global	CHAID
Image1	**527**	649	12172	90.33	**91.57**	82.23	90.95
GLASS	**1950**	2128	2074	71.11	72.60	**73.18**	63.72
BREAST Cancer	3608	**2613**	7784	91.66	91.23	90.05	**93,47**
GREC Radon	**69**	92	2192	90.43	90.17	81.42	**92.94**

References

1. Ganter, B., Kuznetsov, S.: Pattern structures and their projections. In Delugach, H., Stumme, G., eds.: Conceptual Structures: Broadening the Base. Volume 2120 of LNCS. (2001) 129–142
2. Kaytoue, M., Kuznetsov, S.O., Napoli, A., Duplessis, S.: Mining gene expression data with pattern structures in formal concept analysis. Inf. Sci. **181** (2011) 1989–2001
3. Dougherty, J., Kohavi, R., Sahami, M.: Supervised and unsupervised discretization of continuous features. In: Machine Learning: Proc. of the Twelfth International Conference, Morgan Kaufmann (1995) 194–202
4. Muhlenbach, F., Rakotomalala, R.: Discretization of continuous attributes. In Reference, I.G., ed.: Encyclopedia of Data Warehousing and Mining. J. Wang (2005) 397–402
5. Breiman, L., Friedman, J., Olshen, R., Stone, C.: Classification and regression trees. Wadsworth Inc., 358 pp (1984)
6. Kass, G.: An exploratory technique for investigating large quantities of categorical data. Applied Statistics **29(2)** (1980) 119–127
7. Quinlan, J.: Improved use of continuous attributes in C4.5. Journal of Artificial Intelligence Research **4** (1996) 77–90
8. Guillas, S., Bertet, K., Visani, M., Ogier, J.M., Girard, N.: Some links between decision tree and dichotomic lattice. In: Proc. of the Sixth International Conference on Concept Lattices and Their Applications, CLA 2008 (2008) 193–205
9. Ganter, B., Wille, R.: Formal concept analysis, Mathematical foundations. Springer Verlag, Berlin, 284 pp (1999)
10. Fu, H., Fu, H., Njiwoua, P., Nguifo, E.M.: A comparative study of fca-based supervised classification algorithms. In: Concept Lattices. Volume LNCS 2961. (2004) 219–220
11. Birkhoff, G.: Lattice theory. Third edn. Volume 25. American Mathematical Society, 418 pp (1967)
12. Wille, R.: Restructuring lattice theory : an approach based on hierarchies of concepts. Ordered sets (1982) 445–470 I. Rival (ed.), Dordrecht-Boston, Reidel.
13. Girard, N., Bertet, K., Visani, M.: Local discretization of numerical data for galois lattices. In: Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2011 (2011) to appear.
14. Visani, M., Bertet, K., Ogier, J.M.: Navigala: an original symbol classifier based on navigation through a galois lattice. International Journal of Pattern Recognition and Artificial Intelligence, IJPRAI **25** (2011) 449–473