Flexible Concept Lattices

Engelbert Mephu Nguifo
CRIL - IUT de Lens, Rue de l'université SP 16, 62307 Lens cedex. France.
mephu@cril.univ-artois.fr

Keywords: Formal Concept Analysis, Galois lattices, Fuzzy Logic

Galois lattices or Concept lattices is a mathematical framework that builds embedded classes (or concepts) from a context (see figure 1) i.e a set of objects related to a set of attributes [5]. A concept is a node with two parts: an extension and an intension. The extension is the subset of objects which verify the intension, and the intension is the subset of attributes that hold for the extension objects.

O\A	a	b	c	d	e	f	g	h
1	1	1	1	1	1	1	1	
2	1	1	1	1	1	1		1
3	1	1	1	1	1		1	1
4	1	1	1	1		1		
5	1	1		1	1		1	
6	1	1	1		1			1
7	1		1			1		

Figure 1: *A small context $C = (O, A, I)$. If $o_i \in O$ verifies $a_j \in A$, then the box (o_i, a_j) has the value '1'.*

The usefulness of Galois Lattices has been reported in different areas such as data analysis [14], machine learning [2] [9] [11], information retrieval [3] and databases [6].

Many algorithms for non-incremental or incremental determination of Galois lattices were reported [4] [7] [6] [11]. Concept lattices (see figure 2) can grow exponentially in size with respect to their contexts. Even if this happens rarely in practical applications, the lattices exhaustiveness remains a major drawback for the usage of concept lattice-based system, since it considerably increases the time and space complexity of such systems when dealing with large amounts of

data. Applying such lattices-based systems for knowledge discovery in domains with huge amounts of data such as biology or banking, could be problematic.

Whenever the difference between rows or columns is insignificant, the lattices construction would generate many nodes which may be much more similar with fewer effect on the task process. Our intention is to merge such nodes when possible by using fuzzy logic, in order to considerably decrease the computational complexity of lattices-based systems.

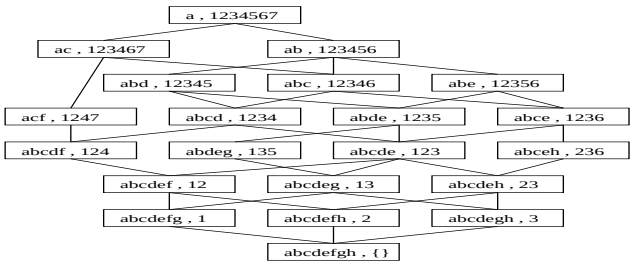

Figure 2: *Galois or Concept lattices.*

For example in figure 1, columns a and b differ only for object 7. This difference appears in the lattices since two nodes are created: (a, 1234567) and (ab, 123456). When the number of objects is huge as it is the case with very large databases, this fact seems to appear very often, even if the difference is very insignificant. To avoid such effect, we propose an approach based on fuzzy logic theory to build flexible concepts in order to approximate the lattices.

Our approach introduces a similarity measure between two attributes such that two attributes will be similar if their corresponding columns in the initial context differ by less than δ objects. δ is a similarity threshold which can be set by the user or automatically computed by the system. This similarity measure allows to build an approximated complete lattices with flexible concepts.

A flexible concept is a concept for which the extension is a fuzzy set [15]. In our case, the extension of a concept consists of two parts: the first part of the set verifies the intension of the concept, whereas the second part doesn't. The second part partially verifies the intension, and it is possible to make the hypothesis that 'the second part may

verify the intension' in case of uncertainty, or to associate to each object, its membership degree.

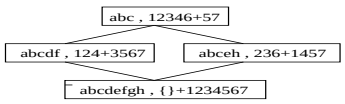

Figure 3: *An approximation of previous lattice.* $\delta = 1$ over the 7 objects.

In the previous example, with $\delta = 1$, columns b and c are similar to column a, so that the lattice node (abc, 12346+57) could be immediately built. Object 5 (respectively 7) may verify the attribute c (resp. b). Different nodes which can be considered as non pertinent are not generated, for example (a, 1234567), (ac, 123467), (ab, 123456). Figure 3 is a flexible concept lattices of figure 2.

Ongoing research is dealing with the formal definition of flexible concept lattices and their properties. Also, different questions arise from such an approach. Once the similarity measure is chosen, is there only one flexible concept lattices? Among all the algorithms designed to build concept lattices, which ones are appropriate to build flexible concepts? Is it possible to introduce a preference relationship among attributes of the context in order to build flexible concepts?

In the area of machine Learning, Mephu & Zegaoui (see [10]) proposed a new algorithm that uses fuzzy logic to generate a learning system based on flexible concept lattices. Preliminary results of experiments show the effectiveness of this approach. When varying δ, results of classification accurary are similar on artificial datasets, whereas on real datasets they increase. At the same time, results of running space and time are very closed on artificial datasets, while on real datasets, they exponentially decrease. This tends to show that this approach doesn't have any effect on artificial data, while on real dataset it is effective. This comes from the fact that real datasets very often contain noisy data, while this is less the case with artificial data.

Different works are being conducted in order to deal with the problem of reduction nodes of concept lattices. Van der Merwe & Kourie (see [13]) proposed a lattices-based data structure called a compressed lattices, which builds the concept lattices before compressing it. Our approach is a way to deal with noisy data or uncertainty among data, and could be a way to build directly a rough sets-based system [12].

Acknowledgments. Financial support has been partially made available by the Ganymede project of the Contrat de Plan Etat/Nord-Pas-de-Calais. Thanks to *C. Zimny* for english proofreading and *O. Roussel* for latex formatting.

References

[1] Bordat J-P. Calcul pratique du treillis de galois d'une correspondance. in *Maths. Sci. Soc journal*, 24ème année, 96:31-47. 1986.

[2] Carpineto C. and Romano G. Galois: An order-theoretic approach to conceptual clustering. in *Proc of ICML'93*. p33-40, Amherst (MS), July 1993.

[3] Carpineto C. and Romano G. Information retrieval through hybrid navigation of lattice representations. in *Intl Journal of Human-Computer Studies*. 45:553-578. 1996.

[4] Ganter B. Two basic algorithms in concept analysis. in *FB4-Preprint nř 831*. TH Darmstadt. 1984.

[5] Ganter B. & Wille R. *Formal Concept Analysis*. Springer-Verlag. 1999.

[6] Godin R. & Missaoui R. An incremental concept formation approach for learning from databases. *Theoretical Computer Science*. 133:387-419. 1994.

[7] Guénoche A. Construction du treillis de Galois d'une relation binaire. in *Math. Sci. Humaines*. 109:41-53. 1990.

[8] Duquenne V. On the core of finite lattices. in *Discrete Math.*. 88:133-147. 1991.

[9] Mephu Nguifo E. Galois Lattice: A framework for Concept Learning. Design, Evaluation and Refinement. in *Proc of the 6th ICTAI'94*. p461-467, New Orleans, Louisiana, November 6-9, 1994. IEEE Press.

[10] Mephu Nguifo E. & Zegaoui A. Classification supervisée à partir de treillis de concepts approximatifs. In *Actes 7èmes Journées de la Société Française de Classification*. 291-298. 1999.

[11] Oosthuizen G.D. Lattice-based Knowledge Discovery. in *Proc. of AAAI-91 KDD workshop*. Annheim, 221-235. 1991.

[12] Pawlak Z. *Rough sets*. Kluwer, Dordrecht-Boston. 1991.

[13] Van der Merwe F.J. & Kourie D.G. A Lattice-Based Data Structure for Information Retrieval and Machine Learning. In *Proc. of ICCS workshop on Concept Lattices for KDD*. Mephu Nguifo & al. Eds. 2001.

[14] Wille R. Restructuring Lattice Theory. In *Proc of Symposium on Ordered Sets*. p445-470. University of calgary, Boston. 1982.

[15] Zadeh L.A. Fuzzy Sets. In *Information and Control*, 8:338-353. 1965.