数理解析研究所講究録 1351 巻 2004 年 122-133

Mining Association Rules using Lattice Theory

Florent Domenach Masato Koda
Institute of Policy and Planning Sciences,
Tsukuba University,
1-1-1 Tenno-Dai,
Tsukuba, Ibaraki 305-8573, Japan.
e-mail : {domenach,koda}@shako.sk.tsukuba.ac.jp

Abstract

The problem of discovering nontrivial association rules from large databases has recently become critical, especially in data mining area where there exists a substantial need to develop efficient mining algorithms for complex data. In this paper, we consider the situation where items are constrained, i.e., some taxonomies (or hierarchies) on the items are known. We first show how taxonomies can be generalized using lattices, which are ordered structures, to represent constraints on items. Then, we propose a new approach to find association rules, based on the notion of biclosures introduced by the first author.

Keywords: Association Rule, Galois Connection, Implication, Lattice.

1 Introduction

Data Mining is essentially done using statistical and computational technics (e.g. principal component analysis, factor analysis, ...), to reveal hidden factors that underlie sets of variables, measurements or signals. In a more algebraic approach, Galois connections first provide the mathematical formalization of the classical extent/intent scheme of objects described by properties (see, e.g., [4, 12, 17]). Given a relation between a set of objects and a set of properties, dually isomorphic orders on classes of objects and sets of shared properties are provided by a natural Galois connection. This basic fact was then frequently rediscovered in the literature. Combined with other considerations, it was successfully developed in Formal Concept Analysis ([16, 17, 30]). It is also the starting point of many studies in Learning

Theory, Conceptual Classification, Relational or Object Databases (see, e.g., [7, 18, 23, 29]).

They have been recognized as a fundamental mathematical concept in the middle of the century ([5, 15, 26]), and constitute a useful tool in several domains related with data analysis like modelization and aggregation of similarities and preferences ([19, 21, 22]) or mathematical morphology ([20]). They are strongly associated to many fundamental notions: among others, closure operators, full implicational systems [2], and, more recently, overhanging relations [10].

In the applications domains mentioned above, there is a need for efficient ways of handling and transforming Galois connections. In that purpose, after recalling some basic definitions in Section 2, we will present here, in a latticial oriented approach, different Data Mining tools, such as implications and their generalization, the association rules (Section 3). Finally, in Section 4, we will introduce what we mean by conceptual classification with constraints.

2 Definitions

First let's recall some basic definitions, together with some well known properties about closure mappings, lattices and Galois connections.

2.1 Closure Mappings and Lattices

Let S be a finite set. A *closure space* is a pair (S, φ), where φ is a *closure operator* on $\mathcal{P}(S)$, that is a mapping onto $\mathcal{P}(S)$ satisfying the following three properties:

(C1) φ is *isotone*: for all $A, B \subseteq S, A \subseteq B$ implies $\varphi(A) \subseteq \varphi(B)$;

(C2) φ is *extensive*: for all $A \subseteq S, A \subseteq \varphi(A)$;

(C3) φ is *idempotent*: for all $A \subseteq S, \varphi(\varphi(A)) = \varphi(A)$.

The image $\mathcal{F}_\varphi = \varphi(\mathcal{P}(S))$ of $\mathcal{P}(S)$ by φ is exactly the set of all the fixed points of φ, which are called the elements of $\mathcal{P}(S)$ *closed* by φ.

A *lattice* is a tuple $(L, \vee, \wedge)$, where L is a set, $x \vee y$ is the lowest upper bound of x and y and $x \wedge y$ is the greatest lower bound of x and y. These two operators are also called respectively *join* and *meet* of x and y (for further information on lattices, see [5, 11]). With the inclusion order, $(\mathcal{F}_\varphi, \vee, \cap)$ is also a lattice with $\vee X = \varphi(\bigcup X)$ for $X \subseteq \mathcal{F}_\varphi$. By the extensivity property, $\varphi(S) = S \in \mathcal{F}_\varphi$.

Example 2.1 A web site selling products can be modelled in the following lattice (Figure 1), the minimum element being the home page, and the maximum the page where the products are sold.

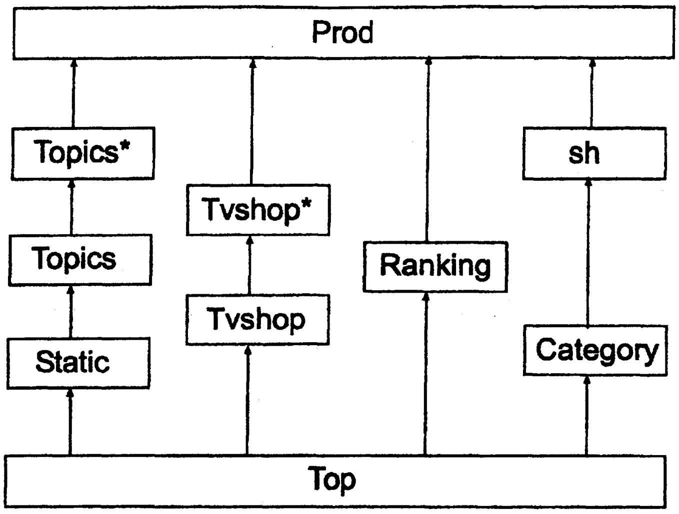

Figure 1: Lattice of a web site.

Example 2.2 Let $\mathcal{H}$ be an hierarchy on S, that is a set of subsets of S (*clusters*) satisfying the conditions: (H1) $S \in \mathcal{H}$, (H2) for $H, H' \in \mathcal{H}, H \cap H' \in \{\emptyset, H, H'\}$, and (H3) for all $s \in S$, $\{s\} \in \mathcal{H}$. The set $\mathcal{H} \cup \{\emptyset\}$ is a closure system on S. Hierarchies constitutes a basic model of classification trees.

Conversely, a closure $\varphi_{\mathcal{F}}$ on $\mathcal{P}(S)$, given by $\varphi_{\mathcal{F}}(F) = \bigcap\{F' \in \mathcal{F} : F \subseteq F'\}$, corresponds, in this way, to any family $\mathcal{F}$ of subsets of S satisfying (i) $S \in \mathcal{F}$, and (ii) $\mathcal{F}' \subseteq \mathcal{F}$ implies $\cap\mathcal{F}' \in \mathcal{F}$. With these properties, $\mathcal{F}$ is said to be a *Moore family* (or a *closure system*) on S, and we denote by $\mathcal{M}$ the set of all Moore families on S. It is well-known that Moore families and closure operators are in one-to-one correspondence and constitute in fact equivalent notions. The set $\mathcal{M}$, ordered by inclusion, is itself a lattice, whose main properties are described in the recent work of [6].

An element j of the lattice $(L, \vee, \wedge)$ is *join irreducible* if $X \subseteq L$ and $j = \vee X$ imply $j \in X$; the set of all join irreducibles of L is denoted as J_L or J if no confusion is possible. The join irreducible elements of L are those which cannot be obtained by others and using the join operator. Dually, an element m of L is *meet irreducible* if $X \subseteq L$ and $m = \wedge X$ imply $m \in X$. The set of all the meet irreducibles of $\mathcal{F}$ is denoted as M_L or M. For an element $x \in L$, the subsets $(x] = \{y \in L : y \leq x\}$ and $[x) = \{y \in L : x \leq y\}$

are respectively called the *principal ideal* and the *principal filter* of L with basis x.

2.2 Galois Mappings

Let L and L' be two complete lattices, and a mapping $f : L \to L'$. A mapping satisfying the following condition is said to be a *Galois mapping* ([26]) :

(GM) *the mapping f is antitone and there exists an antitone mapping g from L' to L such that the composition mappings $\varphi = gf$ and $\psi = fg$ are extensive.*

The pair (f, g) is a *Galois connection* between L and L'; the maps f, g in a Galois connection determine each other uniquely. Both composition mappings φ and ψ are closures, respectively on L and L'. The ordered sets $\Phi = \varphi(L)$ and $\Psi = \psi(L')$ are dually isomorphic by the restrictions of f and g.

The following type of Galois connection is fundamental in data mining. Consider two sets S, A and a relation $R \subseteq S \times A$. For example, S can be a set of objects, A a set of attributes and R is the relation sRa defined by "the object s has the attribute a". For $s \in S, a \in A$, the equivalent notation sRa or $(s, a) \in R$ will be used according to the context. Define $f_R : \mathcal{P}(S) \to \mathcal{P}(A)$ and $g_R : \mathcal{P}(A) \to \mathcal{P}(S)$ by $f_R(C) = \{a \in A : (s, a) \in R \text{ for all } s \in C\}$ and $g_R(D) = \{s \in S : (s, a) \in R \text{ for all } a \in D\}$, for all $C \subseteq S, D \subseteq A$. The mapping f associates to a set of objects C all the attributes shared by all the objects in C, the *intension* of C, and g associates to a set of attributes D all objects having all the attributes in D, the *extension* of D. It is straightforward that the pair (f_R, g_R) satisfies Condition (GM) and constitutes a Galois connection between $\mathcal{P}(S)$ and $\mathcal{P}(A)$, both endowed with the inclusion order. The lattice of closed subsets of S is the *Galois lattice* of R, sometimes also called *(formal) concept lattice.*

Example 2.3 Set $P = \mathcal{P}(S)$, where S is the finite set of objects under study, and consider a (complete) lattice Q of *descriptions*, together with a description $d(s) \in Q$ of each element $s \in S$. The order of Q corresponds with a generalization order, where $q \leq q'$ means that description q is more general than description q' (an example of such a lattice is the one of Example 2.1). Then, it is said that $s \in S$ satisfies description q if $q \leq d(s)$. For any class (subset) $C \subseteq S$, and description $q \in Q$, the mappings $f(C) = \bigwedge\{d(s) : s \in$

$C\}$ and $g(q) = \{s \in S : q \leq d(s)\}$ constitutes a Galois connection between $\mathcal{P}(S)$ and Q.

Remark 2.1 Example 2.3 can appear to be a generalization of the previous type of Galois connection; in fact, as shown in [17], the Formal Concept Lattice scheme is the more general case (see [9] for a presentation of a common frame).

The well-known Galois connection associated with the *table* of finite lattice T belongs to the previous type. The table of T is the relation $R_T \subseteq J \times M$ defined by $(j, m) \in R_T$ if $j \leq m$, for any $j \in J, m \in M$. Every finite lattice is isomorphic to the Galois lattice associated with its table ([4]). The generalization of this fact to any complete lattice was given by [30].

2.3 Galois Connections in Data Mining

One basic purpose of any Data Mining is to obtain classes of objects sharing similar characters, a description by attributes being associated to each class. As described above, Galois connections, by providing a correspondence between extents and intents, satisfy this purpose.

Another important aim in Data Mining is to organize data to make it more readable or to recover some unknown structure. For instance, hierarchical clustering methods provide a classification tree, sometimes (e.g. in phylogenetic reconstruction or cognitive psychology) an estimation of an unknown tree. The Galois lattice does not correspond to such an objective since it preserves the whole information of the data. So, as frequently observed in the literature, it has a great sensitivity to noise and deviation from the model. Also, the number of concepts potentially grows exponentially with the data size, leading to problems of computational complexity.

Between many different approach, one can pick methods to prune the concept lattice, for instance by limiting its construction to a convenient filter [24]. Another approach, more practically oriented, is to consider weakened conditions for the closure systems associated, or for the (equivalent) full implicational systems. We will present in the following Section such an approach.

3 Lattices and Conceptual Classification

3.1 Implications

Full implicational systems constitute a notion equivalent with both closure operators and Moore families. An *implicational system* on S is a binary relation $\mathcal{S} \subseteq \mathcal{P}(S) \times \mathcal{P}(S)$ on $\mathcal{P}(S)$. In the sequel, $(A, B) \in \mathcal{S}$ is denoted $A \to_{\mathcal{S}} B$ (or $A \to B$ if no confusion is possible). We then say that A *implies* B or that $A \to B$ is an *implication* (of $\mathcal{S}$).

A *full implicational system* (frequently abbreviated as CIS) is an implicational system satisfying the following conditions:

(S1) $B \subseteq A$ implies $A \to B$;

(S2) for any $A, B, C \in S, A \to B$ and $B \to C$ imply $A \to C$ (transitivity);

(S3) for any $A, B, C, D \in S, A \to B$ and $C \to D$ imply $A \cup C \to B \cup D$.

[2] has established a one-to-one correspondence between closure spaces (or closure systems) and full implicational systems. First, given a closure operator φ on S, the implicational system $\mathcal{S}_\varphi = \{A \to B : B \subseteq \varphi(A)\}$ is full. Conversely, if $\mathcal{S}$ is a full implicational system on S, then the set $\mathcal{F}_{\mathcal{S}} = \{F \subseteq S : X \subseteq F \text{ and } X \to Y \text{ imply } Y \subseteq F\}$ is a closure system on S. Its associated closure operator is denoted as $\varphi_{\mathcal{S}}$. When $\mathcal{F}_{\mathcal{S}}$ is a classification scheme, the meaning of $A \to B$ is that any class containing the elements of A contains also those of B. Some examples of studies using implicational systems were made, beyond many others, by Duquenne ([13, 14]) and Diday ([7]).

3.2 Association Rules

Full implicational systems satisfies strong requirements that, at a first glance, could be expected to be rarely satisfied, or practically useless. So a possible extension of the notion of implication can be found in Agrawal and al ([1, 27]), with the definition of *association rules.* An association rule consists of two itemsets (called the antecedent and the consequent), denoted by $A \to B$, with $A \cap B = \emptyset$. The *support* of an association rule is the number of items satisfying $A \cup B$, and the *confidence* is the probability with which the items in A appear together with the items in B in the given dataset. More, we have:

$$conf(A \to B) = \frac{sup(A \cup B)}{sup(A)}$$

For example, an association rule $A \rightarrow B$ have a confidence 0.9 if 90% of the items supporting A also support B.

Remark 3.1 *An implication is an association rule with a confidence of 1, justifying our use of it as a generalization of implications.*

4 Constrained Conceptual Classification

In this section, we will present a particular type of binary relations, called *biclosed relations*, first introduced in [8]. This type of relation is particularly interesting to study because it is in a one-to-one correspondence with Galois connection. So, instead of dealing with mappings, we may work on biclosed relations. Moreover, it can be use to modelize constraints on items or on properties.

4.1 Biclosed relations on a product of closure spaces

In this paragraph, we introduce a type of binary relations on $S \times S'$, called *biclosed relations*. This type of binary relation is in a one-to-one correspondence with Galois connections, and so, instead of having a couple of mappings (constituting the Galois connection), it's often more easy to use binary relations. All the missing proofs can be found in [8].

Let (S, φ) and (S', φ') be two closure spaces, with the corresponding Moore families Φ and Φ', respectively on S and S'. A relation $R \subseteq S \times S'$ is said *biclosed* if it satisfies the following conditions:

(B1) for any $a \in S, aR = \{a' \in S' : (a, a') \in R\} \in \Phi'$;

(B2) for any $a' \in S', Ra' = \{a \in S : (a, a') \in R\} \in \Phi$;

Condition (B1) corresponds to the closure on rows, while (B2) corresponds in the same way to the closure on columns.

The set of all the biclosed relations is denoted as $\mathcal{R}_{\varphi\varphi'}$. The closure on $\mathcal{P}(S \times S')$ associated with the Moore family $\mathcal{R}_{\varphi\varphi'}$ is denoted as Γ. So $\Gamma(R)$ is the intersection of all the biclosed relations containing R. Consider the following two mappings $\Gamma_1, \Gamma_2 : \mathcal{P}(S \times S') \rightarrow \mathcal{P}(S \times S')$ defined, for $R \subseteq S \times S'$, by:

$$\Gamma_1(R) = \{(a, a') \in S \times S' : a' \in \varphi'(aR)\} \text{ and}$$

$$\Gamma_2(R) = \{(a, a') \in S \times S' : a \in \varphi(Ra')\}$$

These two mappings correspond to the two previous conditions (B1) and (B2), i.e. Γ_1 is associated with (B1), for the rows closure, and Γ_2 is associated with (B2) and the columns closure. There is a relationship between these two mappings and Γ, the closure associated with the Moore family $\mathcal{R}_{\varphi\varphi'}$, given by:

Proposition 4.1 *There exists an integer* $k \leq\mid S \times S' \mid$ *such that* $(\Gamma_1\Gamma_2)^k(R) = \Gamma(R)$.

Denoting by $\Phi \otimes \Phi'$ the set of all Galois mappings from Φ to Φ', we have:

Theorem 4.1 *The sets* $\mathcal{R}_{\varphi\varphi'}$ *and* $\Phi \otimes \Phi'$ *are order isomorphic.*

The previous results are in fact valid for S and S' finite or not. When S and S' are finite, we are in a much more simple case. Considering the sets of join irreducibles $J = J_S$ and $J' = J_{S'}$, they are minimal sup-generating sets of S and S' respectively. We have, by Theorem 4.1, an isomorphism between the Galois mappings from S to S' and the biclosed relations between J and J'. Moreover, the biclosed relation R associated to a Galois mapping f is given by

$$jRj' \iff j' \leq f(j)$$

for all $j \in J, j' \in J'$. Conversely, the Galois mapping f associated to a biclosed relation R is the classical one f_R, as defined in Section 2.2.

Finally, it will be equivalent in many cases to consider biclosed relations and Galois connections. This isomorphism is particularly interesting to consider, because instead of dealing with mappings (Galois connections), we can now work on binary relations.

4.2 Using Lattices as Constraints

In practice, some knowledge is often present before extracting any implications or association rules. Or users are interested in a subset of implications, or of association rules. For example, they may only want rules that contain a specific item or rules that contain children of a specific item in a lattice. This *a priori* knowledge can easily be expressed in terms of subsets of a lattice or of a full implicational system.

The first type of constraints is consisted of a given lattice, different from the boolean lattice $\mathcal{P}(S)$ (the lattice of all subsets of S, which is the lattice without any constraint). One can see the web site lattice in Example 2.1. The most frequent case is to have a taxonomy on the items; however, as seen in Example 2.2, taxonomies (*is-a* hierarchies) are just particular cases of closure systems. In that case, to obtain rules containing a specific item x, we just have to consider the principal ideal $(x]$ from the lattice, and use it as constraint (it's itself a lattice). Obtaining rules containing a set $O = \{x_1, \ldots, x_n\}$ of objects is equivalent, in the same way, as the conjunction of the principal ideals $(x_i]$.

The other type of constraints, more frequently founded, is an a priori set $\mathcal{C}$ of implications given by an expert. In fact, as seen in Section 3.1, full implicational systems and lattices are in a one-to-one correspondence, and so this type of constraints is just a particular case of the previous type. From this set $\mathcal{C}$ of implications, we will prune the boolean lattice to obtain a particular lattice satisfying all the implications in $\mathcal{C}$.

So, as soon as constraints may be described by a closure (or a lattice, or a set of implications), as well on the set of items and/or on the set of properties, we can use an algorithm to "biclose" the data relatively to these constraints, for example using the algorithm described in [8], and only after extract implications or association rules, using already known algorithms [1, 27].

5 Conclusion

The understanding of both association rules and biclosed relation can lead us to develop an new algorithm extracting such rules in a data table having constraints. Such a work had been done by [3, 28], but only using taxonomies as constraints, and only on the set of items. The theory developed in [8] allows us to extent these results to any type of lattice, and, more, to put constraints on both items and properties.

References

[1] R. Agrawal and T. Imielinski and A. Swami (1993), Mining Associations between Sets of Items in Massive Databases, *Proc. of the ACM-SIGMOD 1993 Int'l Conference on Management of Data*, Washington D.C., May 1993, 207-216.

[2] W.W. Armstrong (1974), Dependency structures of data base relationships, *Information Processing* 74, 580–583.

[3] R.J. Bayardo Jr., R. Agrawal and D. Gunopulos (2000), Constraint-Based Rule Mining in Large, Dense Databases. *Data Mining and Knowledge Discovery Journal*, 4 (2/3).

[4] M. Barbut, B. Monjardet (1970), *Ordre et classification, Algebre et combinatoire*, Paris, Hachette.

[5] G. Birkhoff (1967), *Lattice Theory* (3rd ed.), Providence, R.I., Amer. Math.Soc.

[6] N. Caspard and B. Monjardet (2003), The lattices of Moore families and closure operators on a finite set: a survey, *Discrete Applied Math.*, 127 (2), 241–269.

[7] E. Diday (1988), The symbolic approach in clustering and related methods of data analysis: the basic choices, *Classification and Related Methods of Data Analysis*, H.H. Bock, Amsterdam, North-Holland, 673-683.

[8] F. Domenach and B. Leclerc (2001), Biclosed binary relations and Galois connections, *Order* 18, 89–104.

[9] F. Domenach and B. Leclerc (2002), On the Roles of Galois Connections in Classification. In: Schwaiger, M., Opitz, O., (Eds.), *Explanatory Data Analysis in Empirical Research.* Studies in Classification, Data Analysis and Knowledge Organization, Invited lecture at the 25th Conference of the German Classification Society (Munich, march 2001). Munich, Springer-Verlag, 31-40.

[10] F. Domenach and B. Leclerc (2003), Closure systems, Implicational Systems, Overhanging Relations and the case of Hierarchical Classification. *Mathematical Social Sciences*, to appear.

[11] B.A. Davey and H.A. Priestley (2002), *Introduction to lattices and order* (2nd ed.), Cambridge University Press.

[12] V. Duquenne (1987), Contextual implications between attributes and some representation properties for finite lattices. In Ganter et al. (Eds.), Beiträge zur Begriffsanalyse, Wissenchaftverlag, Mannheim, 213–240.

[13] V. Duquenne (1995), Models of possessions and lattice analysis, *Social sciences information* 34 (2), 253-267.

[14] V. Duquenne (1996), Lattice analysis and the representation of handicap associations, *Social Networks* 18, 217-230.

[15] C.J. Everett (1944), Closure Operators and Galois Theory in Lattices, *Trans. Amer. Math. Soc.* 55, 514-525.

[16] B. Ganter, R. Wille (1998), Appendix H. Applied lattice theory: formal concept analysis, in G. Grätzer, *General Lattice Theory* (2nd edition), Basel, Birkhauser Verlag, 591-605.

[17] B. Ganter, R. Wille (1999), *Formal concept analysis*, Springer.

[18] R. Godin, R. Missaoui, H. Alaoui (1991), Learning algorithms using a Galois lattice structure, *Proceedings of the 1991 IEEE Int. Conf. for AI*, San Jose, CA, 22-29.

[19] M.F. Janowitz (1978), An order theoretic model for cluster analysis, *SIAM J. Appl. Math.* 37, 55-72.

[20] R. Keshet (2000), Mathematical morphology on complete semilattices and its applications to image processing, *Fundamenta Informaticae* 41, 33-56.

[21] B. Leclerc (1991), Aggregation of fuzzy preferences: a theoretic Arrow-like approach, *Fuzzy sets and systems* 43 (3), pp. 291-309.

[22] B. Leclerc (1994), The residuation model for the ordinal construction of dissimilarities and other valued objects, in B. Van Cutsem, *Classification and Dissimilarity Analysis*, New York, Springer-Verlag, 149-172.

[23] M. Liquière, J. Sallantin (1998), Structural machine learning with Galois lattices and graphs, *International Conference on Machine Learning, ICML98*, 1-9.

[24] E. Mephu Nguifo (1993), Une nouvelle approche basée sur le treillis de Galois pour l'apprentissage de concepts. *Math. Inf. Sci. hum.*, Vol. 124, 19–63.

[25] Mirkin, B.: On the problem of reconciling partitions. In: Quantitative Sociology, International Perspectives on Mathematical and Statistical Modelling. New-York, Academic Press, 441-449 (1975).

[26] O. Öre (1944), Galois connections, *Trans. Amer. Math. Soc.* 55, 494-513.

[27] R. Srikant and R. Agrawal (1997), Mining Generalized Association Rules, *Future Generation Computer Systems* 13(2-3).

[28] R. Srikant, Q. Vu and R. Agrawal (1997), Mining Association Rules with Item Constraints. In: Proc. of the 3rd Int'l Conference on Knowledge Discovery in Databases and Data Mining, Newport Beach, California.

[29] K. Waiyamai, R. Taouil, L. Lakhal (1998), Towards an object database approach for managing concept lattice based applications, in *16th Int. Conf. on Conceptual Modeling (ER'97), Los Angeles, CA*, Lectures notes in Computer Science 1331, Berlin, Springer-Verlag, 299-312.

[30] R. Wille (1982), Restructuring lattice theory: an approach based on hierarchies of concepts, in I. Rival, *Ordered Sets*, Dordrecht, D. Reidel, 445-470.