Transformations of Concept Graphs: An Approach to Empirical Induction *

John L. Pfaltz

Dept. of Computer Science, Univ. of Virginia
jlp@virginia.edu

1 Concept Lattices

Let R be a binary relation, as illustrated in Figure 1(a), with m rows and n columns. Such a relation can represent many phenomena and there exists an extensive literature on relational algebras [1]. In this paper we take a more limited view and simply regard R as an observation of a set of attributes A associated with a set of objects O. In our formulation, objects are denoted by numbered rows and attributes denoted by lettered columns.

"Formal Concept Analysis" [5] has been developed by Rudolf Wille [15], Bernard Ganter and their colleagues at Darmstadt. In their approach the construction and visual display of *concept lattices*, that is partially ordered sets of concepts, is crucial. The nodes of the concept lattice correspond to abstract *concepts* of the phenomenon being modelled and relationships within the lattice are reflective of relationships in the external world. Their book has numerous examples, and their method has found application in industrial applications (reported by Ganter & Wille) and in code re-engineering [8, 14].

What makes it interesting to this workshop is an investigation of how these concept lattices are transformed with the advent of new information.

We begin with a very brief overview of concept lattices. Let R be a binary relation between any two sets O and A, as in Figure 1(a). We regard O as a set of *objects* and A as a set of *attributes*. But, they can be arbitrary sets. For example, Lindig and Snelting [8] apply concept analysis to legacy code by creating a relation R between P, a set of procedures, and V, a set of global variables.

By the closure, φ_R of O with respect to R, we mean a maximal set of objects which share the same attributes as all $o \in O$. Similarly, $\varphi_{R^{-1}}$ operating on a set A of attributes picks up any other attributes that are common to all objects which satisfy each $a \in A$.[2] Ganter and Wille [5] show that φ_R and $\varphi_{R^{-1}}$ are indeed closure operators, and constitute a Galois connection. For any R, such as that of Figure 1(a), the closure systems of φ_R and $\varphi_{R^{-1}}$ are isomorphic and can be represented by the lattice $\mathcal{L}_R$ of closed sets shown in Figure 1(b), which are partially ordered by inclusion. Labeling each node is the pair of closed sets that is joined by the Galois connection, for example $< abg, 123 >$.

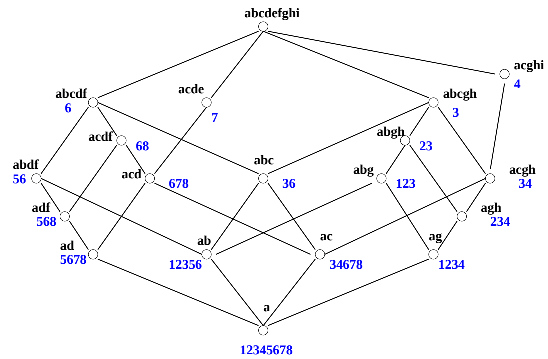
* Research supported in part by DOE grant DE-FG05-95ER25254.

[2] More formally, the Galois closure, φ_R, on O with respect to R consists of those closed sets $\bar{O} \subseteq O$ of the form $\bar{O} = O_i.\overline{R}.\overline{R^{-1}}$, for $O_i \subseteq O$, where $O_i.\overline{R} = \bigcap_{o \in O_i} o.R \subseteq A$ and $A_i.\overline{R^{-1}} = \bigcap_{a \in A_i} a.R^{-1} \subseteq O$. Conversely, one forms the closure, $\varphi_{R^{-1}}$ of A with respect to R consisting of the closed sets $\bar{A} = A_k.\overline{R^{-1}}.\overline{R}$. The set $O_i.\overline{R}$ denotes the set of all attributes shared by every object in O. Consequently, $O_i.\varphi = O_i.\overline{R}.\overline{R^{-1}}$ denotes the set of *all* the objects that share (at least) these common attributes. Similarly, $A_k.\overline{R^{-1}}$ denotes the set of all objects sharing every attribute in A_k and $A_k.\varphi = A_k.\overline{R^{-1}}.\overline{R}$ consists of *all* the attributes shared by the objects which (at least) have Y in common.

A

O	a	b	c	d	e	f	g	h	i
1	×	×					×		
2	×	×					×	×	
3	×	×	×				×	×	
4	×		×				×	×	×
5	×	×		×		×			
6	×	×	×	×		×			
7	×		×	×	×				
8	×		×	×		×			

(a)



(b)

Fig. 1. A relation R (a) and its concept lattice $\mathcal{L}_R$ (b)

The set *abg* is closed in A; 123 is closed in O. In this case we have oriented the lattice with respect to A, the set of attributes, where the universe $A = abcdefghi$ (which must be closed) is the lattice *supremum*. The singleton set $\{a\}$, which is an attribute of every object is the lattice $infimum$. It is partially ordered with respect to set inclusion.

Readily, the concept lattice $\mathcal{L}$ is a visual model of the content of R. There are many similar examples of applied concept analysis in Ganter's and Wille's book [5]. Later extensions to concept analysis are reported in [16].

2 Closure Spaces

An operator φ is a *closure operator* if $X \subseteq X.\varphi$, $X \subseteq Y \Rightarrow X.\varphi \subseteq Y.\varphi$, and $X.\varphi.\varphi \subseteq X.\varphi$. The Galois closure on binary relations is one kind of discrete closure operator. A more general treatment of *closure spaces* has been advanced in [9, 12]. A central idea in these papers is that of the *generators* of a closed set, Z, denoted $Z.\gamma$, by which we mean a minimal set Y such that $Y.\varphi = Z$. For example, with a convex hull closure operator, the generators of a convex n-gon are its n vertices (or extreme points).[3]

An n-gon is uniquely determined by its generators. Whenever the generators of a closed set must be unique, we say the closure operator is uniquely generated and call the resulting closure space an *antimatroid*.[4] Much of the closure literature, *e.g.* [2, 3, 4, 9, 12] assumes antimatroid closure.

Using concepts from closure spaces, it is quite straightforward to generate the concept lattice while simultaneously determining the generators of these closed concepts. For example, the single attribute e generates the closed concept *acde*. That is, $\{e\}.\varphi_{R^{-1}} = \{acde\}$. To see this in R, observe that every object which has property e (there is only one!) also has properties a, c and d as well. Similarly we find that either $\{bd\}$ or $\{bf\}$ will generate $\{abdf\}$ because attributes b and d only found together in objects 5 and 6, which also share attributes *abdf*. This closure space, and most

[3] In the discrete geometry literature [3] all generators are called *extreme* points.

[4] Matroids and antimatroids are identical, except that the closure operator of a matroid satisfies an exchange axiom while the closure operators of an antimatroid satisfies an anti-exchange axiom.

arising from concept analysis, are not antimatroid. Nevertheless, they retain much of the structure of antimatroid closure spaces [7].

If we regard R as a relation in the database sense, then $(o, a) \in R$ denotes that a is an attribute of object o. In Figure 1(a) it is clear that $abgh$ are shared attributes of objects 2 and 3. Attributes bh generate $abgh$. So we may assert that *in this world* $(\forall o \in O)[o.bh \Rightarrow o.abgh]$; or more simply we have the attribute implication $bh \Rightarrow abgh$ Similarly one may show that both bcd and bcf are minimal generators of $abcdf$; so we have the attribute implication $bcd \vee bcf \Rightarrow abcdf$. By deriving the generators of all the closed concept sets, we extract all the logical implications (universally quantified over O) that are valid for R. From now on we will use $\Rightarrow$ to denote both attribute implication and closure generation.

The interpretation of $X.\gamma$ and $X.\varphi$ as precedent and consequent respectively in a rule based description of a discrete world opens up a entire new approach to knowledge discovery [11] that can be exploited in relatively small discrete worlds.[5] Although this cursory description of the generators of closed sets in a concept lattices may be too brief for full comprehension, it should be sufficient to suggest the potential for modeling inductive learning by incrementally adding observations (rows) to R.

3 Inductive Transformations

If a concept lattice $\mathcal{L}_R$ captures all the logical attribute implications one can make about a collection of objects;[6] it is natural to ask "suppose we observe one more object and its attributes. How will this transform the lattice $\mathcal{L}_R$?" This is the essence of discrete, empirical induction. Given a collection R of observations that have an internal structure denoted by $\mathcal{L}_R$, how does new information transform this structure? Actually, such transformations are inherently "graceful" and "local" in nature because of a fundamental property of closed sets — the intersection of closed sets must be closed. This leads to an interesting interplay between closed sets $Z = X.\varphi$ and their generators $Z.\gamma$.

Every time we add a row (object/attributes observation) to R, we add at least one new closed set to $\mathcal{L}_R$, because the attributes of a single row constitute a closed set of A.[7] Let $< o', A' >$ denote this new row. If there exists $Z \in \mathcal{L}_R$ such that $A' = Z$, then the lattice remains unchanged. Suppose not. Then, there exists at least one closed Z in $\mathcal{L}_R$ such that $A' \subset Z$. We consider $A' \cap Y$ for all closed $Y, Y \subset Z$. These are the only elements of the concept lattice $\mathcal{L}_R$ with which A' can interact.

For example, appending to R a new observation of object 9 with attributes a, c and g yields the relation of Figure 2(a) and the corresponding concept lattice $\mathcal{L}_{R_2}$ of Figure 2(b). Observe that $acg \subset acgh$ and $acg \cap agh = ag$. This local interaction occurs in the lower right corner, where a single new concept (closed set) acg has been added yielding new relationships that are indicated by dashed lines.

This newly observed datum has also changed the generation structure of $\mathcal{L}_R$. In $\mathcal{L}_R$, we have $(cg \vee ch) \Rightarrow acgh$. In $\mathcal{L}_{R_2}$, we have $cg \Rightarrow acg$, so cg can no longer be a generator of $argh$. Now in $\mathcal{L}_{R_2}$, $ch \Rightarrow acgh$.

[5] A robot project at U.Va. [6] gathers sensor data about objects in its world in a relational table. It will use our algorithm to convert this data into implications for input to its rule based planning component.

[6] In [5], R_1 was obtained by assertions about pond life made in a child's educational TV show. It is literally a child-like understanding of real phenomena.

[7] This need not be strictly true; but it is typical. Further, *wlog* we may assume it.

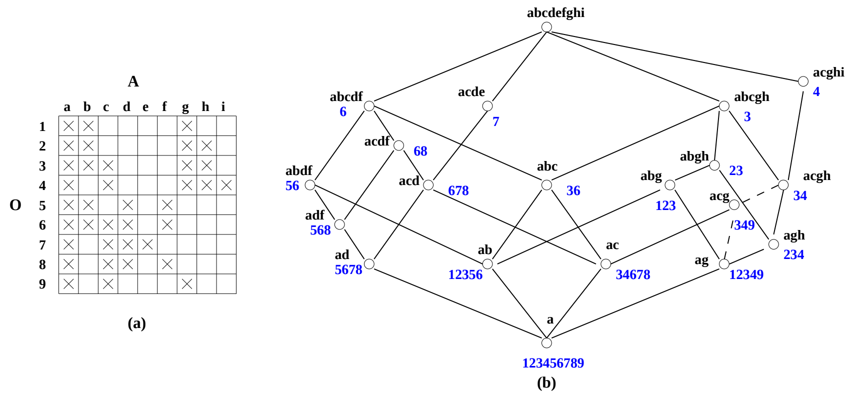


Fig. 2. R_2 and its concept lattice $\mathcal{L}_{R_2}$

We observe that this new object is not very different from existing objects. It is contained in $Z = acgh$, which is fairly low in $\mathcal{L}_R$. Suppose $Z = abcdefghi = \mathbf{U}$, the universe of attributes? One can show that ef is a generator of $abcdefghi$, along with 11 other minimal generators. But, there are no objects associated with $abcdefghi = A$. In this world, ef is a logical contradiction.[8] In Figure 3(a) we have now changed the new object 9 so it has the attributes a, d, e and f. The combination ef is no longer a contradiction. In $\mathcal{L}_{R_3}$, $adef$ is covered by $Z = \mathbf{U}$. It intersects *acde* and *abcdf* (which are also covered by Z) in *ade* and *adf* respectively. The closed set *ade* is new, and it recursively intersects with *acd* (which is also covered by *acde*) as *ad*. The changes in Figure 3(b)

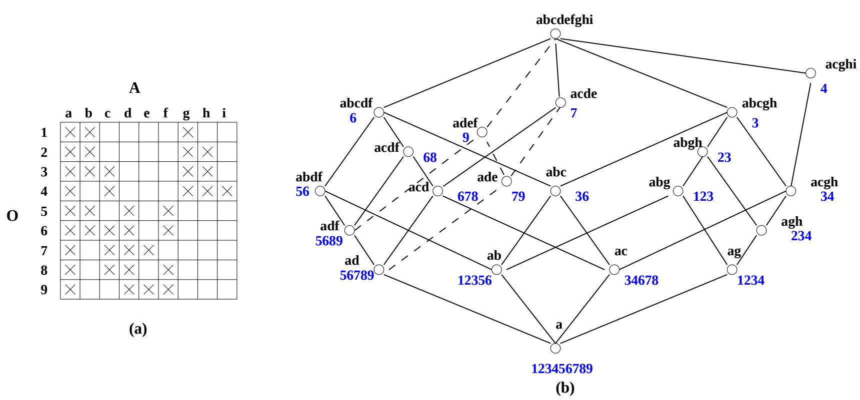


Fig. 3. R_3 and its concept lattice $\mathcal{L}_{R_3}$

are again indicated by dashed lines.

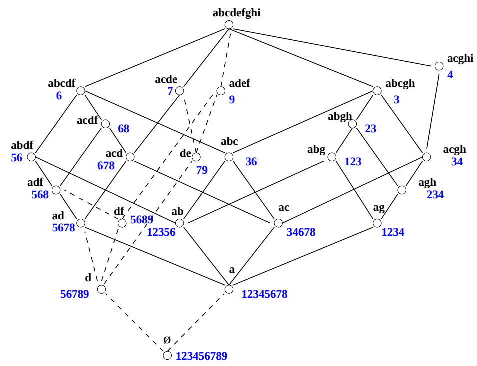
[8] One of the strengths of this approach to knowledge discovery is that in addition to deriving all true implications, it also identifies all logical contradictions which cannot be true in *this world of objects*.

For a final example, we observe that a is an attribute of every object. It corresponds to logical tautology in the universe of R. By adding a 9^{th} row with only attributes def we change that. It intersects with *acde* and *abcdf* to create *de* and *df* respectively and the interesting concept lattice of Figure 4(b).

A

O	a	b	c	d	e	f	g	h	i
1	×	×					×		
2	×	×					×	×	
3	×	×	×				×	×	
4	×		×				×	×	×
5	×	×		×		×			
6	×	×	×	×		×			
7	×		×	×	×				
8	×		×	×		×			
9				×	×	×			

(a)



(b)

Fig. 4. R_4 and its concept lattice $\mathcal{L}_{R_4}$

Addition of new rows (empirical observations) to the logical world described by a binary relation R engenders a regular graceful transformation of the concept lattice based on iterated set intersection. Conversely, it has been shown [10, 13] that *deletion* of an element from an antimatroid closure space induces a lattice homomorphism τ on its closure lattice $\mathcal{L}$.[9] As observed earlier, concept closure spaces are not normally antimatroid. We conjecture, but have not yet proven, that deletion in concept lattices will still induce at least a meet homomorphism.

Together, these results would indicate that the gradual accumulation of "knowledge" based on sequential, empirical observation is relatively "stable". Certainly, this is in accord with our intuitive, psychological understanding of knowledge. But, this is still very active research. For example, we conjecture that as the concept lattice becomes large, the expected magnitude of incremental change will become small. Also, we would like to know what a major restructuring of the concept lattice (a world understanding) would look like — and what might cause it.

References

1. Chris Brink, Wolfram Kahl, and Gunther Schmidt. *Relational Methods in Computer Science*. Springer Verlag, Wien, 1997.
2. Paul H. Edelman. Meet-Distributive Lattices and the Anti-exchange Closure. *Algebra Universalis*, 10(3):290–299, 1980.
3. Paul H. Edelman and Robert E. Jamison. The Theory of Convex Geometries. *Geometriae Dedicata*, 19(3):247–270, Dec. 1985.

[9] Actually τ is not quite a lattice homomorphism. It is meet preserving, therefore order preserving. But, it only preserves the *supremum* of a set if the *supremum* covers the set.

4. Martin Farber and Robert E. Jamison. Convexity in Graphs and Hypergraphs. *SIAM J. Algebra and Discrete Methods*, 7(3):433–444, July 1986.
5. Bernard Ganter and Rudolf Wille. *Formal Concept Analysis - Mathematical Foundations.* Springer Verlag, Heidelberg, 1999.
6. James P. Gunderson and Worthy N. Martin. The Effects of Uncertainity on Plan Success in a Simulated Maintenance Robot Domain. *J. of Experimental and Theoretic AI*, 12:153–164, 2000.
7. Robert E. Jamison and John L. Pfaltz. Closure Spaces that are not Uniquely Generated. In *Ordinal and Symbolic Data Analysis, OSDA 2000*, Brussels, Belgium, July 2000.
8. Christian Lindig and Gregor Snelting. Assessing Modular Structure of Legacy Code Based on Mathematical Concept Analysis. In *Proc of 1997 International Conf. on Software Engineering*, pages 349–359, Boston, MA, May 1997.
9. John L. Pfaltz. Closure Lattices. *Discrete Mathematics*, 154:217–236, 1996.
10. John L. Pfaltz. Deletion Transformations in Antimatroid Closure Spaces. In *25th Combinatoric, Computing, and Graph Theory Conf.*, Boca Raton, FL, Mar. 1998.
11. John L. Pfaltz. Closure Spaces and Knowledge Discovery. In *Princ. Knowledge Disc. and Data Mining, PKDD 2001*, Freiburg, Germany, 2001. (Submitted).
12. John L. Pfaltz and Robert E. Jamison. Closure Systems and their Structure. In Jules Desharnais, editor, *5th Intern'l Seminar on Relational Methods in Computer Science, RelMICS 2000*, pages 121–132, Valcartier, Quebec, Jan. 2000.
13. John L. Pfaltz and John E. Karro. Transformations of Antimatroid Closure Spaces. Technical Report TR CS-98-13, Dept of Computer Sci., University of Virginia, June 1998.
14. Gregor Snelting and Frank Tip. Reengineering Class Hierarchies Using Concept Analysis. In *Proc. ACM SIGSOFT 6th International Symposium on Foundations of Software Engineering, FSE-6*, pages 99–110, Lake Buena Vista, FL, 1998.
15. Rudolf Wille. Restructuring lattice theory: An approach based on hierarchies of concepts. *Ordered Sets*, pages 445–470, 1982.
16. Karl Erich Wolff. Concepts, States, and Systems. In *CASYS'99, 3rd Conf. on Computing Anticipatory Systems*, Liege, Belgium, Aug. 1999.

This article was processed using the LaTeX macro package with LLNCS style