Theoretical Foundations of Association Rules

Mohammed J. Zaki * and Mitsunori Ogihara †
Computer Science Department, University of Rochester, Rochester NY 14627
{zaki,ogihara}@cs.rochester.edu

Abstract

In this paper we describe a formal framework for the problem of mining association rules. The theoretical foundation is based on the field of formal concept analysis. A concept is composed of closed subsets of attributes (itemsets) and objects (transactions). We show that all frequent itemsets are uniquely determined by the frequent concepts. We further show how this lattice-theoretic framework can be used to find a small rule generating set, from which one can infer all other association rules.

1 Introduction

Association rule discovery, a successful and important mining task, aims at uncovering all frequent patterns among sets (or transactions) composed of data attributes. Most of the current work has focused on developing efficient algorithms [2, 3, 4, 15, 19, 20, 23, 24, 25, 29]. On the other hand, there has been little work in formulating a theory of associations. Such a theory can help in estimating the complexity of the mining task, and also in developing a unified framework for common data mining problems.

This paper begins by presenting some complexity results based on the connection between frequent itemsets and bipartite cliques. We then place association rule mining within the lattice-theoretic framework of formal concept analysis introduced by Wille [28]. Given a binary relation, a *concept* consists of an *extent* (transactions) and an *intent* (attributes), such that all objects in the extent share the attributes in the intent, and vice versa. We show that all frequent itemsets are uniquely determined by the set of *frequent concepts*. We then tackle the problem of generating a *base*, a minimal rule set, from which all the other association rules can be inferred. The concept lattice framework can not only aid in the development of efficient algorithms, but can also help in the visualization of discovered associations, and can provide a unifying framework for reasoning about associations, and supervised (classification) and unsupervised (clustering) concept learning [5, 6, 10].

The rest of the paper is organized as follows. We present the association rule problem statement in Section 2. A graph-theoretic view of the problem is given in Section 3. Section 4 casts association mining as a search for frequent concepts, and Section 5 looks at the problem of generating rule bases. We discuss related work in Section 6 and conclude in Section 7.

2 Association Rules: Problem Formulation

The association mining task, introduced in [2], can be stated as follows: Let $\mathcal{A}$ be a set of items, and $\mathcal{T}$ a database of transactions, where each transaction has a unique identifier (*tid*) and contains a set of items. A set of items is also called an *itemset*. The *support* of an itemset X, denoted $\sigma(X)$, is the number of transactions in which it occurs as a subset. An itemset is *frequent* if its support is more than a user-specified *minimum support (min_sup)* value. An *association rule* is an expression $A \rightarrow B$, where A and B are itemsets. The support of the rule is given as $\sigma(A \cup B)$, and the *confidence* as $\sigma(A \cup B)/\sigma(A)$ (i.e., the conditional probability that a transaction contains B, given that it contains A). The mining task consists of two steps [3]: 1) Find all frequent itemsets. This step computationally and

*Supported in part by the NSF grants CCR-9409120 and CCR-9705594, ARPA contract F19628-94-C-0057, and U.S. Air Force/Rome Labs contract F30602-95-1-0025.

†Supported in part by the NSF grants CCR-9701911, CCR-9725021, and INT-9726724.

I/O intensive. 2) Generate high confidence rules. This step is relatively straightforward; rules of the form $X \backslash Y \rightarrow Y$ (where $Y \subset X$), are generated for all frequent itemsets X, provided the rules have at least *minimum confidence (min_conf)*. Figure 1a shows a bookstore database with six customers who buy books by different authors. Figure 1b shows all the frequent itemsets with $min\_sup = 50\%$ (i.e., those occurring in at least 3 transactions), and the set of all association rules with $min\_conf = 80\%$. $ACTW$ and CDW are the maximal-by-inclusion frequent itemsets. Since all other frequent itemsets are subsets of these two itemsets, we can reduce the mining problem to the enumeration of only the maximal frequent itemsets.

ITEMS

Jane Austen	A
Agatha Christie	C
Sir Arthur Conan Doyle	D
Mark Twain	T
P. G. Wodehouse	W

DATABASE

Transcation	Items
1	A C T W
2	C D W
3	A C T W
4	A C D W
5	A C D T W
6	C D T

FREQUENT ITEMSETS (min_sup = 50%)

Support	Itemsets
100% (6)	C
83% (5)	W, CW
67% (4)	A, D, T, AC, AW CD, CT, ACW
50% (3)	AT, DW, TW, ACT, ATW CDW, CTW, ACTW

Maximal Frequent Itemsets: CDW, ACTW

Association Rules with Confidence = 100%

A → C (4/4)	AC → W (4/4)	TW → C (3/3)
A → W (4/4)	AT → C (3/3)	AT → CW (3/3)
A → CW (4/4)	AT → W (3/3)	TW → AC (3/3)
D → C (4/4)	AW → C (4/4)	ACT → W (3/3)
T → C (4/4)	DW → C (3/3)	ATW → C (3/3)
W → C (5/5)	TW → A (3/3)	CTW → A (3/3)

Association Rules with Confidence >= 80%

W → A (4/5)	C → W (5/6)	W → AC (4/5)

Figure 1: a) Example Database, b) Frequent Itemsets and Rules with $min\_conf = 80\%$

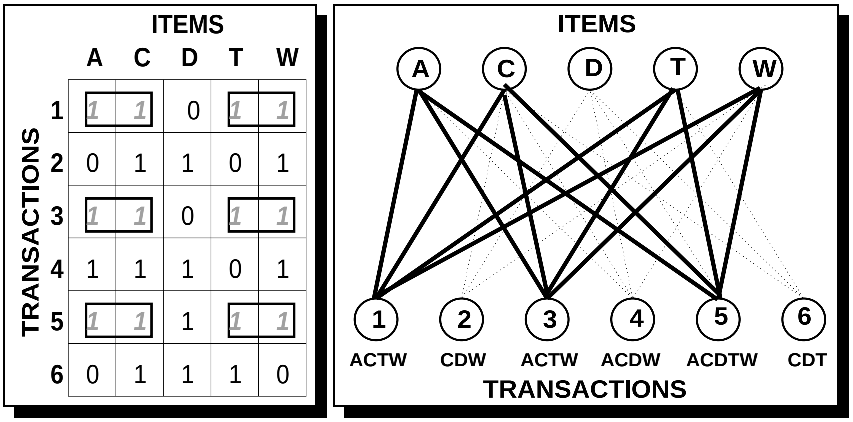
3 Itemset Discovery: Bipartite Graphs

Definition 1 *A* **bipartite graph** $G = (U, V, E)$ *has two distinct vertex sets* U *and* V*, and an edge set* $E = \{(u, v) \mid u \in U \text{and } v \in V\}$*. A complete bipartite subgraph is called a* **bipartite clique**.

Definition 2 *A* **hypergraph** *on* $\mathcal{I}$ *is a family* $H = \{E_1, E_2, ..., E_n\}$ *of edges or subsets of* $\mathcal{I}$*, such that* $E_i \neq \emptyset$*, and* $\cup_{i=1}^{n} E_i = \mathcal{I}$*. A set* $T \subset \mathcal{I}$ *is a* **transversal** *of* H *if it intersects all the edges, that is to say:* $T \cap E_i \neq \emptyset \ \ \forall E_i$.

The input database for association mining is essentially a very large bipartite graph, with U as the set of items, V as the set of tids, and each (item, tid) pair as an edge. The problem of enumerating all (maximal) frequent itemsets corresponds to the task of enumerating all (maximal) constrained bipartite cliques, $I \times T$, where $I \subseteq U$, $T \subseteq V$, and $|T| \geq min\_sup$. Due to the one-to-one correspondence between bipartite graphs, binary matrices and hypergraphs, one can also view it as the problem of enumerating all (maximal) unit submatrices in a binary matrix (see Figure 2a), or as the problem of enumerating all (minimal) transversals of a hypergraph, satisfying the support constrains (a transversal of a hypergraph is the complement of an independent set, which in turn is a clique in the complimentary hypergraph). Figure 2b shows the bipartite graph of the database and the maximal constrained bipartite clique $ACTW \times 135$ (the maximal frequent itemset $ACTW$).

Figure 2c shows the complexity of decision problems for maximal bipartite cliques (itemsets) with restrictions on the size of I (items) and T (support). For example, the problem whether there exists a maximal bipartite clique such that $I + T \geq K$ (with constant K) is in P, the class of problems that can be solved in polynomial time. On the other had, the problem whether there exists a maximal bipartite clique such that $I + T = K$ is NP-Complete [17], the class of "hard" problems for which no polynomial time algorithm is known to exist. The last row of the table may seem contradictory. While there is unlikely to exist a polynomial time algorithm for finding a clique with $I + T \leq K$,



	<=K	=K	>=K
|I|	P	NP-Complete	P
|T|	P	NP-Complete	P
|I|+|T|	NP-Complete	NP-Complete	P

Figure 2: a) Maximal Unit Submatrix, b) Maximal Constrained Bipartite Clique $ACTW \times 135$, c) Mining Complexity

the largest cliques with $I + T \geq K$ can be found by reducing it to the maximum matching problem [16], which has $O((U + V)^{2.5})$ complexity. The following theorem says that counting the number of maximal cliques in a bipartite graph is extremely hard.

Theorem 1 ([17]) *Determining the number of maximal bipartite cliques in a bipartite graph is #P-Complete.*

The complexity results shown above are quite pessimistic, and apply to general bipartite graphs. We should therefore focus on special cases where we can find polynomial time solutions. Fortunately for association mining, in practice the bipartite graph (database) is very sparse, and we can in fact obtain linear complexity in the graph size.

The *arboricity* $r(G)$ of a graph is the minimum number of forests into which the edges of G can be partitioned, and is given as $r(G) = \max_{H \subset G} \{e(H)/(n(H) - 1)\}$, where $n(H)$ is the number vertices and $e(H)$ the number of edges of the subgraph H. A bound on the arboricity is equivalent to a notion of hereditary sparsity. For a bipartite graph $r(G) = I \cdot T/(I + T - 1)$, where $I \times T$ is a maximum bipartite clique. Furthermore, if we assume $I \ll T$ (as is generally the case in practice, since we want large support), then $r(G) \approx I$, i.e., the arboricity is given by the maximum sized frequent itemset. For sparse graphs, of bounded arboricity I, the complexity of finding all maximal bipartite cliques is linear in number of items and transactions:

Theorem 2 ([8]) *All maximal bipartite cliques can be enumerated in time $O(I^3 \cdot 2^{2I} \cdot (U + V))$.*

Even though the above algorithm has linear complexity, it is not practical for large databases due to the large constant overhead (I can easily be around 10 to 20 in practice). Nevertheless, the result is very encouraging, and provides the reason why all current association mining algorithms exhibit linear scalability in database size. This result also says that at least in theory the association mining algorithms should scale linearly in the number of items or attributes, a very important feature if practicable.

Theorem 3 ([16]) *All maximum independent sets can be listed in $O((U + V)^{2.5} + \gamma)$ time, where γ is the output size.*

The above theorem states that all the maximum (largest) bipartite cliques (independent sets in complimentary graph) of a bipartite graph can be found in time polynomial in input, and linear in the output size. However, due to the greater than quadratic complexity, it remains to be seen if this algorithm is practical for large databases with millions of transactions.

4 Itemset Discovery: Formal Concept Analysis

In this section we will show that association mining is very closely related to *formal concept analysis*, which was introduced in a seminal paper by Wille [28]. We assume that the reader is familiar with basic concepts of lattice theory (see [7] for a good introduction).

Definition 3 *Let S be a set. A function $c : \mathcal{P}(S) \mapsto \mathcal{P}(S)$ defined between the powerset of S, is a* **closure operator** *on S if, for all $X, Y \subseteq S$, c satisfies the following properties:*

1) Extension: $X \subseteq c(X)$.

2) Monotonicity: if $X \subseteq Y$, then $c(X) \subseteq c(Y)$.

3) Idempotency: $c(c(X)) = c(X)$.

A subset X of S is called **closed** *if $c(X) = X$.*

Definition 4 *A* **context** *is a triple (G, M, I), where G and M are sets and $I \subseteq G \times M$. The elements of G are called* **objects***, and the elements of M are called* **attributes***. For an arbitrary $g \in G$, and $m \in M$, we note gIm, when g is related to m, i.e., $(g, m) \in I$.*

Definition 5 *Let (G, M, I) be a context with $X \subseteq G$, and $Y \subseteq M$. Then the mappings*

$$s : G \mapsto M, s(X) = \{m \in M \mid \ (\forall g \in X) \ gIm\}$$

$$t : M \mapsto G, t(Y) = \{g \in G \mid \ (\forall m \in Y) \ gIm\}$$

define a **Galois connection** *between $\mathcal{P}(G)$ and $\mathcal{P}(M)$, the power sets of G and M, respectively.*

The set $s(X)$ is the set of attributes common to all the objects in X and $t(Y)$ is the set of objects common to all the attributes in Y. We note that $X_1 \subseteq X_2 \Rightarrow s(X_2) \subseteq s(X_1)$, for $X_1, X_2 \subseteq G$ (and "dually" for function t on M). Furthermore, the compositions $c = s \circ t$ and dually, $t \circ s$ are closure operators.

Definition 6 *A* **concept** *of the context (G, M, I) is defined as a pair (X, Y), where $X \subseteq G$, $Y \subseteq M$, $s(X) = Y$, and $t(Y) = X$. In other words, a concept (X, Y) consists of the closed sets X and Y, since $X = t(Y) = t(s(X)) = s \circ t(X) = c(X)$, and similarly $Y = c(Y)$. X is also called the* **extent** *and Y the* **intent** *of the concept (X, Y).*

The concept generated by a single attribute $m \in M$ given as $\alpha(m) = (t(m), c(m))$ is called an *attribute concept*, while the concept generated by a single object $g \in G$ given as $\beta(g) = (c(g), s(g))$ is called an *object concept*.

The set of all concepts of the context is denoted by $\mathcal{B}(G, M, I)$. A concept (X_1, Y_1) is a *subconcept* of (X_2, Y_2), denoted as $(X_1, Y_1) \leq (X_2, Y_2)$, iff $X_1 \subseteq X_2$ (iff $Y_2 \subseteq Y_1$). Notice that the mappings between the closed sets of G and M are anti-isomorphic, i.e., concepts with large extents have small intents, and vice versa. The different concepts can be organized as a hierarchy of concepts based on the superconcept-subconcept partial order.

Definition 7 *A subset P of an ordered set Q is* **join-dense** *if $\forall q \in Q$, there exists $Z \subseteq P$, such that $q = \bigvee_Q Z$ (and dually we can define* **meet-dense***).*

Theorem 4 ([28]) **Fundamental Theorem of Formal Concept Analysis:** *Let (G, M, I) be a context. Then $\mathcal{B}(G, M, I)$ is a complete lattice with join and meet given by*

$$\bigvee_j (X_j, Y_j) = (c(\bigcup_j X_j), \bigcap_j Y_j) \qquad \bigwedge_j (X_j, Y_j) = (\bigcap_j X_j, c(\bigcup_j Y_j))$$

Conversely, if L is a complete lattice then L is isomorphic to $\mathcal{B}(G, M, I)$ iff there are mappings $\gamma : G \mapsto L$, and $\mu : M \mapsto L$, such that $\gamma(G)$ is join-dense in L, $\mu(M)$ is meet-dense in L, and gIm is equivalent to $\gamma(g) \leq \mu(M)$ for all $g \in G$ and $m \in M$. In particular L is isomorphic to $\mathcal{B}(L, L, \leq)$.

The complete lattice $\mathcal{B}(G, M, I)$ is called the *Galois* lattice of the context. The concept lattice can be represented graphically by a *Hasse diagram*, where each concept is a circle, and for concepts $c_1 \leq c_2$, there is a line joining them, with c_1 being lower than c_2. For example, Figure 3a shows the Galois lattice for our example database. It is shown with a *minimal labeling*, where the intent (extent) of a concept can be reconstructed by considering all labels reachable above (below) that concept. In other words, each concept is labeled with an attribute (object) if it is an attribute (object) concept. It is clear that an appropriately drawn diagram can aid in visualizing and understanding the relationships among the attributes and objects (i.e., associations).

Define a *frequent concept* as a concept (X, Y) with $X \subseteq G$, $Y \subseteq M$, and $|X| \geq min\_sup$. Figure 3b shows all the frequent concepts with *min_sup* = 50%. All frequent itemsets can be determined by the meet operation on attribute concepts. For example, since meet of attribute concepts D and T, $\alpha(D) \wedge \alpha(T)$, doesn't exist, DT is not frequent, while $\alpha(A) \wedge \alpha(T) = (135, ACTW)$, thus AT is frequent. Furthermore, the support of AT is given by the cardinality of the resulting concept's extent, i.e., $\sigma(AT) = |135| = 3$. Thus all frequent itemsets are uniquely determined by the frequent concepts. This observation can possibly aid the development of efficient algorithms since we need to enumerate only the closed frequent itemsets, instead of enumerating all frequent itemsets like most current algorithms.

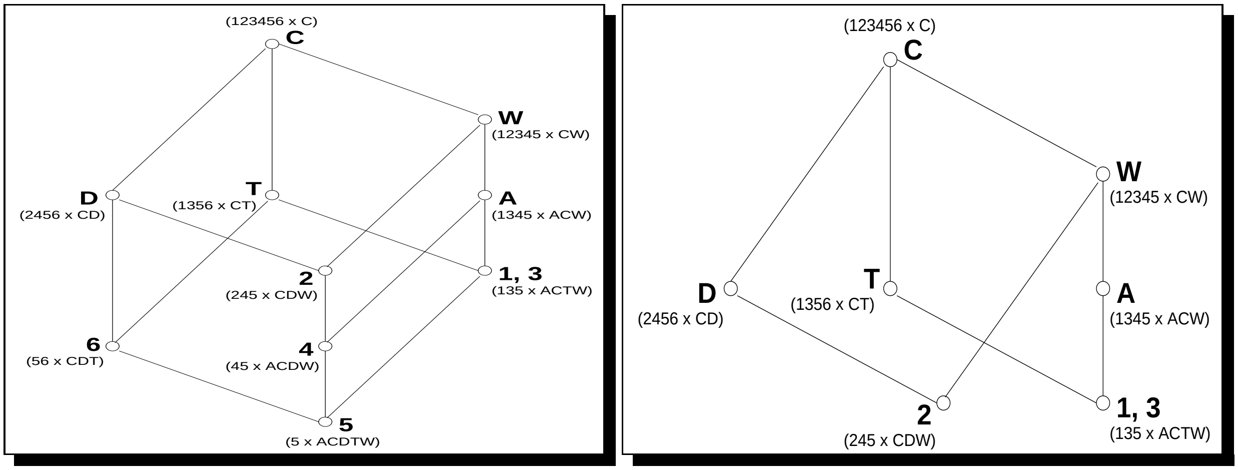


Figure 3: a) Galois Lattice of Concepts; b) Frequent Concepts

5 Rule Generation

Association rules were originally proposed in [2]. However, we will show below that association rules are exactly the *partial implications*, satisfying support and confidence constraints, proposed in an earlier paper [21].

Let (G, M, I) be a context. A *partial implication rule* $X \xrightarrow{p} Y$ is a triple (X, Y, p), where $X, Y \subseteq M$ are sets of attributes, and the *precision* $p = P(Y|X) = |t(X \cup Y)|/|t(X)|$. Association rules correspond to partial implications meeting the support and confidence constraints, i.e., with $|t(X \cup Y)| \geq min\_sup$, and $p \geq min\_conf$, respectively.

Let $\mathcal{I} = \{X \xrightarrow{p} Y \mid X, Y \subseteq M, p = P(Y|X)\}$, be a set of partial implications. A rule $K \xrightarrow{p} L$ can be *derived* from $\mathcal{I}$, denoted $\mathcal{I} \vdash K \xrightarrow{p} L$, iff we can obtain $K \xrightarrow{p} L$ from $\mathcal{I}$ by applying certain inference rules. In this case we also call $K \xrightarrow{p} L$ a *redundant* rule according to $\mathcal{I}$.

Definition 8 *A set $\mathcal{R} \subseteq \mathcal{I}$ is called a* **generating set** *for $\mathcal{I}$ iff $\mathcal{R} \vdash \mathcal{I}$. A minimal generating set is called a* **base***.*

Since the set of all partial implications (i.e., association rules) can be very large, we are interested in finding a base for it. This means that only a small and easily understandable set of rules can be presented to the user, who can later selectively derive other rules of interest. The set of partial implications can be broken into two parts: implications with $p = 1$ and with $p < 1$. A base for all partial implications can be obtained by combining the bases of these two sets.

Global Implications A *global implication rule* is denoted as $X \rightarrow Y$, where $X, Y \subseteq M$, and $t(X) \subseteq t(Y)$, i.e., all objects related to X are also related to Y. It can be shown that $t(X) \subseteq t(Y) \Leftrightarrow P(Y|X) = 1$. Thus, global implications are precisely the association rules with 100% confidence. A global implication can be directly discerned from the Hasse diagram of the concept lattice, since in this case the meet of attribute concepts in X is less than (lies below) the meet of attribute concepts in Y. For example, consider the frequent concepts in Figure 3b. $AD \rightarrow CW$, since $\alpha(A) \wedge \alpha(D) = (45, ACDW) \leq \alpha(C) \wedge \alpha(W) = (12345, CW)$. The problem of finding a base of all global implication rules has been well studied [9, 12, 22, 27]. One characterization of a base is given as follows:

Theorem 5 ([9]) *The set $\{X \rightarrow c(X) \backslash X \mid X$ is a pseudo-intent$\}$ is a base for all global implications, where X is a* **pseudo-intent** *if $X \neq c(X)$, and for all pseudo-intents $Q \subset X$, $c(Q) \subseteq X$.*

For example, $\{A, D, T, W, CTW\}$ is the set of pseudo-intents in our example database. A base of global implications is thus given by the set $\mathcal{R} = \{A \rightarrow CW,\ D \rightarrow C,\ T \rightarrow C,\ W \rightarrow C,\ CTW \rightarrow A\}$. All other global implications can be derived from $\mathcal{R}$ by application of simple inference rules such as those given in [22, pp. 47], 1) Reflexivity: $X \subseteq Y$ implies $Y \rightarrow X$, 2) Augmentation: $X \rightarrow Y$ implies $XZ \rightarrow YZ$, 3) Transitivity: $X \rightarrow Y$ and $Y \rightarrow Z$ implies $X \rightarrow Z$, and so on.

Proper Partial Implications We now turn to the problem of finding a base for *proper* partial implications with $p < 1$, i.e., association rules with confidence less than 100%. Note that for any $Z \subseteq X$, $P(Y|X) = P(Y \cup Z|X)$, and thus $X \xrightarrow{p} Y$ iff $X \xrightarrow{p} Y \cup Z$. In particular, $X \xrightarrow{p} Y$ iff $X \xrightarrow{p} X \cup Y$. We thus only discuss the rules $X \xrightarrow{p} Y$, with $X \subseteq Y$. Furthermore, it can be shown that $X \xrightarrow{p} Y$ iff $c(X) \xrightarrow{p} c(Y)$. We can thus restrict ourselves to only the rules where X and Y are intents of a frequent concept. The set of all proper partial implications is given by $\mathcal{I}^{<1}(\mathcal{B}(G, M, I)) = \{K \xrightarrow{p} L \mid K \subset L$ are intents of $\mathcal{B}(G, M, I)\}$. The following theorem states that unlike global implications, partial implications satisfy transitivity and commutativity only under certain conditions.

Theorem 6 ([21]) *Let $M_1, M_2, M_3, M_4 \subseteq M$ be intents with $M_1 \subseteq M_2 \subseteq M_4$ and $M_1 \subseteq M_3 \subseteq M_4$. Then $P(M_2|M_1) \cdot P(M_4|M_2) = P(M_4|M_1) = P(M_3|M_1) \cdot P(M_4|M_3)$ (i.e., $M_1 \xrightarrow{p} M_2$ and $M_2 \xrightarrow{q} M_4$ implies $M_1 \xrightarrow{pq} M_4$).*

Consider the Hasse diagram of the frequent concepts with the precision on the edges, shown in Figure 4. The edge between attribute concepts C and W corresponds to the implication $C \xrightarrow{5/6} W$. The reverse implication $W \rightarrow C$ has precision 1 by definition. Only the implications between adjacent concepts need to be considered, since the other implications can be derived from the above theorem. For example, $C \rightarrow A$ has precision $p = 4/6$, since $P(A|C) = P(W|C) \cdot P(A|W) = 5/6 \cdot 4/5 = 4/6$. The diagram provides a wealth of embedded information; the link joining attribute concept T and object concept $1, 3$ corresponds to the rule $T \rightarrow A$. Immediately we can see that it has the same confidence $(3/4)$ as the rules $T \rightarrow W,\ T \rightarrow AC,\ T \rightarrow AW,\ T \rightarrow ACW,\ CT \rightarrow A,\ CT \rightarrow W,$ and $CT \rightarrow AW$. All these other rules are thus redundant! On the other hand the link from A to $1, 3$ corresponds to the rule $A \rightarrow T$, which generates another set of redundant rules.

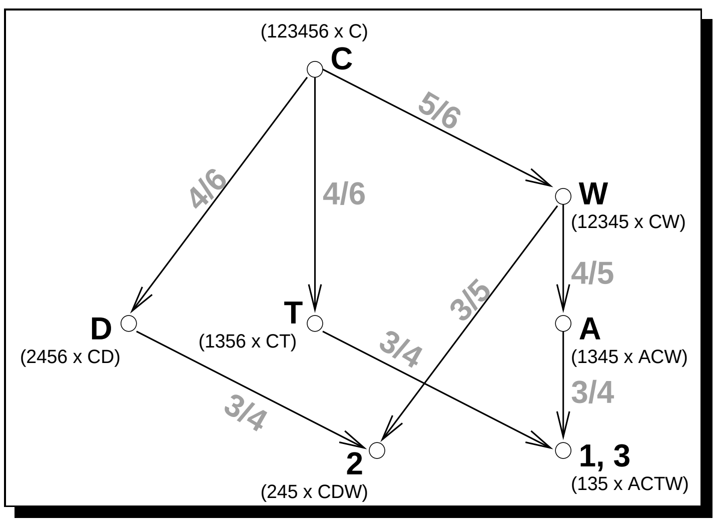


Figure 4: Frequent Concepts with Edge Precisions

For $\mathcal{I}' \subseteq \mathcal{I}^{<1}(\mathcal{B}(G, M, I))$, define the graph $\mathcal{G}(\mathcal{I}') = (V, E)$, with vertex set $V = \{N \subseteq M \mid N$ is an intent$\}$, and edge set $E = \{(K, L) \in V \times V \mid K \xrightarrow{p} L \in \mathcal{I}'\}$.

Lemma 1 ([21]) *If there exists a cycle in $\mathcal{G}(\mathcal{I}')$, then there exists a partial implication $K \in \mathcal{I}$ such that $\mathcal{I} \backslash K \vdash K$.*

As a consequence of this lemma, one rule in every cycle is redundant, and it can be discarded. The next theorem gives a more precise characterization of a generating set.

Theorem 7 ([21]) *$\mathcal{I}'$ is a generating set if 1) $Gr(\mathcal{I}')$ is a spanning tree. 2) M is a consequent of only one partial implication in $\mathcal{I}'$.*

Figure 5a shows a generating set (a minimal spanning tree) for all the proper partial implications in our example. We can derive the precision of a redundant rule by multiplying the precisions of the other rules involved in the cycle (except, we need to invert the precision if we go from a lower concept to a higher concept in the cycle). For example, the precision of the missing edge $D \rightarrow W$ can be obtained by multiplying the inverted precision on the edge from D to C, with the precisions on the edges from C to W, and from W to 2, i.e., $\frac{6}{4} \cdot \frac{5}{6} \cdot \frac{3}{5} = \frac{3}{4}$.

To obtain the rules satisfying a given value of *min_conf*, one can simply discard all edges in the diagram with $p < min\_conf$ For example, Figure 5b shows the generating set for rules with $min\_conf = 80\%$.

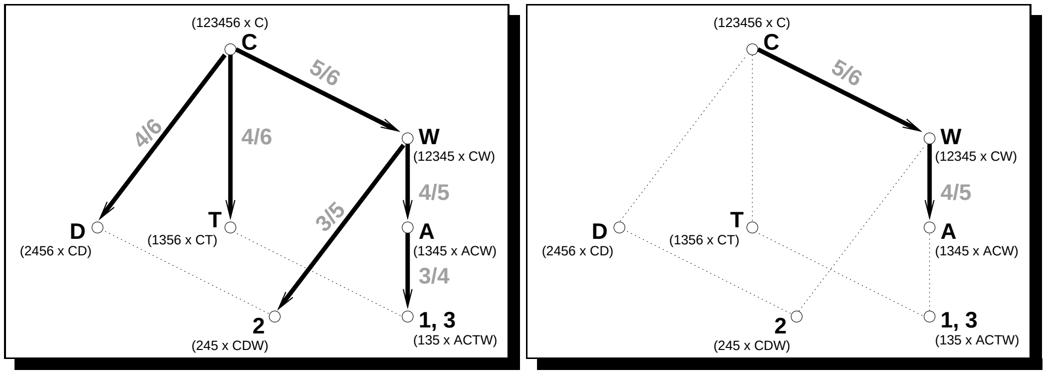


Figure 5: a) Generating Set for $p < 1$; b) Generating Set for $0.8 \leq p < 1$

Definition 9 *An element $x \in L$ of the lattice L, is called* **join-irreducible** *(dually* **meet-irreducible***) if it has exactly one lower (dually upper) neighbor.*

Let $\mathcal{J}(L)$ and $\mathcal{M}(L)$ denote the set of all join- and meet-irreducible elements, respectively. Any finite lattice $(L, \leq)$ is uniquely determined (up to isomorphism) by $\mathcal{J}(L)$ and $\mathcal{M}(L)$, and restricting the order relation to the set $\mathcal{J}(L) \cup \mathcal{M}(L)$. For the lattice in Figure 3a, $\mathcal{J}(L) = \{2, 4, 6, 13\}$, and $\mathcal{M}(L) = \{A, D, T, W\}$. This observation can help in reducing the size of the original database. The following theorem gives upper and lower bounds on the size of a base.

Theorem 8 ([21]) *If $\mathcal{I}' \subseteq \mathcal{I}^{<1}(\mathcal{K} = \mathcal{B}(G, M, I))$ is a base, then* $1/2 \cdot |\mathcal{J}(\mathcal{K}) \cap \mathcal{M}(\mathcal{K})| \;\; \leq \;\; |\mathcal{I}'| \;\; \leq \;\; |\mathcal{K}| - 1$

The bad news is that the upper bound is tight for a large number of lattices, and thus for such concept lattices the construction of a base will not lead to a reduction in storage over a generating set. Furthermore, the lower limit is not very interesting since there exist lattices with $\mathcal{J}(\mathcal{K}) \cap \mathcal{M}(\mathcal{K}) = \emptyset$. The problem of finding a canonical base for all partial implications is thus open. Nevertheless, the generating set obtained by the application of Theorem 7 should be a good substitute for a base in practice. For example, by combining the base for rules with $p = 1$ and the generating set for rules with $p \geq 0.8$, we obtain a generating set for all association rules with $min\_sup = 50\%$, and $min\_conf = 80\%$: $\{A \xrightarrow{1} CW,\ D \xrightarrow{1} C,\ T \xrightarrow{1} C,\ W \xrightarrow{1} C,\ CTW \xrightarrow{1} A,\ C \xrightarrow{5/6} W,\ W \xrightarrow{4/5} A\}$. It can be easily verified that all the association rules shown in Figure 1b can be derived from this set.

6 Related Work

There has been an astonishing amount of research in developing efficient algorithms for mining frequent itemsets [1, 2, 3, 4, 15, 19, 20, 23, 24, 25, 29]. In [14, 13], the connection between associations and hypergraph transversals was made. They also presented a model of association mining as the discovery of maximal elements of theories, and gave some complexity bounds.

A lot of algorithms have been proposed for generating the Galois lattice of concepts [5, 9, 10, 11, 18]. An incremental approach for building the concepts was studied in [6, 10]. These algorithms will have to be adapted to enumerate only the frequent concepts. Further, they have only been studied on small datasets. It remains to be seen how scalable these approaches are compared to the association mining algorithms. Finally, there has been some work in pruning discovered association rules by forming rule covers [26]. However, the problem of constructing a base or generating set has not been studied previously.

7 Conclusions

In this paper we presented a lattice-theoretic foundation for the task of mining associations based on formal concept analysis. We showed that the set of frequent concepts uniquely determines all the frequent itemsets. The lattice of frequent concepts can also be used to obtain a rule generating set from which all associations can be derived. We showed that while there exists a characterization of a base for rules with 100% confidence, the problem of constructing a base for all associations is still open.

References

[1] C. Aggarwal and P. Yu. Online generation of association rules. In *14th Intl. Conf. on Data Engineering*, February 1998.

[2] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *ACM SIGMOD Conf. Management of Data*, May 1993.

[3] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. Inkeri Verkamo. Fast discovery of association rules. In U. Fayyad and et al, editors, *Advances in Knowledge Discovery and Data Mining*, pages 307–328. AAAI Press, Menlo Park, CA, 1996.

[4] S. Brin, R. Motwani, J. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. In *ACM SIGMOD Conf. Management of Data*, May 1997.

[5] C. Carpineto and G. Romano. Galois: An order-theoretic approach to conceptual clustering. In *10th Intl. Conf. on Machine Learning*, June 1993.

[6] C. Carpineto and G. Romano. A lattice conceptual clustering system and its application to browsing retrieval. *Machine Learning*, 24:95–112, 1996.

[7] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 1990.

[8] D. Eppstein. Arboricity and bipartite subgraph listing algorithms. *Information Processing Letters*, 51:207–211, 1994.

[9] B. Ganter. Algorithmen zur formalen begriffsanalyse. *Beiträge zur Begriffsanalyse (Ganter, Wille, Wolff, eds.)*, Wissenschaft-Verlag, 1987.

[10] R. Godin, R. Missaoui, and H. Alaoui. Incremental concept formation algorithms based on galois (concept) lattices. *Computational Intelligence*, 11(2):246–267, 1991.

[11] A. Guenoche. Construction du treillis de galois d'une relation binaire. *Math. Sci. Hum.*, 109:41–53, 1990.

[12] J. L. Guigues and V. Duquenne. Familles minimales d'implications informatives resultant d'un tableau de donnees binaires. *Math. Sci. hum.*, 24(95):5–18, 1986.

[13] D. Gunopulos, R. Khardon, H. Mannila, and H. Toivonen. Data mining, hypergraph transversals, and machine learning. In *16th ACM Symp. Principles of Database Systems*, May 1997.

[14] D. Gunopulos, H. Mannila, and S. Saluja. Discovering all the most specific sentences by randomized algorithms. In *Intl. Conf. on Database Theory*, January 1997.

[15] M. Houtsma and A. Swami. Set-oriented mining of association rules in relational databases. In *11th Intl. Conf. Data Engineering*, 1995.

[16] T. Kashiwabara, S. Masuda, K. Nakajima, and T. Fujisawa. Generation of maximum independent sets of a bipartite graph and maximum cliques of a circular-arc graph. *J. of Algorithms*, 13:161–174, 1992.

[17] S. O. Kuznetsov. Interpretation on graphs and complexity characteristics of a search for specific patterns. *Nauchn. Tekh. Inf., Ser. 2 (Automatic Document. Math. Linguist.)*, 23(1):23–37, 1989.

[18] S. O. Kuznetsov. Fast algorithm for construction of all intersections of the objects from a finite semi-lattice. *Nauchn. Tekh. Inf., Ser. 2 (Automatic Document. Math. Linguist.)*, 9:11–21, 1993.

[19] D-I. Lin and Z. M. Kedem. Pincer-search: A new algorithm for discovering the maximum frequent set. In *6th Intl. Conf. Extending Database Technology*, March 1998.

[20] J-L. Lin and M. H. Dunham. Mining association rules: Anti-skew algorithms. In *14th Intl. Conf. on Data Engineering*, February 1998.

[21] M. Luxenburger. Implications partielles dans un contexte. *Math. Inf. Sci. hum.*, 29(113):35–55, 1991.

[22] D. Maier. *The Theory of Relational Databases*. Computer Science Press, 1983.

[23] J. S. Park, M. Chen, and P. S. Yu. An effective hash based algorithm for mining association rules. In *ACM SIGMOD Intl. Conf. Management of Data*, May 1995.

[24] A. Savasere, E. Omiecinski, and S. Navathe. An efficient algorithm for mining association rules in large databases. In *21st VLDB Conf.*, 1995.

[25] H. Toivonen. Sampling large databases for association rules. In *22nd VLDB Conf.*, 1996.

[26] H. Toivonen, M. Klemettinen, P. Ronkainen, K. Hätönen, and H. Mannila. Pruning and grouping discovered association rules. In *MLnet Wkshp. on Statistics, Machine Learning, and Discovery in Databases*, April 1995.

[27] M. Wild. Implicational bases for finite closure systems. *Arbeitstagung, Begriffsanalyse und Künstliche Intelligenz (W. Lex, ed.)*, pages 147–169, 1991.

[28] R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. *Ordered Sets (I. Rival, ed.)*, pages 445–470, Reidel, Dordrecht-boston, 1982.

[29] M. J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li. New algorithms for fast discovery of association rules. In *3rd Intl. Conf. on Knowledge Discovery and Data Mining*, August 1997.