Query Learning and Certificates in Lattices

Marta Arias   José L. Balcázar

LARCA Research Group
Technical University of Catalonia, Barcelona
{marias,balqui}@lsi.upc.edu

October 15, 2008

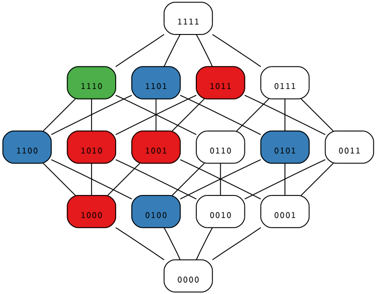
Introduction

- **Why?** Better understand the AFP Algorithm
  - Angluin, Frazier, Pitt: **Learning Conjunctions of Horn Clauses** Machine Learning 9 (1992)
- **Our contributions:** New way at looking at it
  - **AFP-L**:
    - Generalization to abstract, lattice-theoretic setting
    - Simpler proof
  - **Certificates**:
    - Improved construction that applies to strong proper learning
    - Link between AFP learning algorithm and certificates

Preliminaries, I

Horn logic

- Syntax: propositional **variables** are $\{a, b, c, d\}$; with variables we construct **clauses** $ac \rightarrow d$; Horn CNFs are conjunctions of clauses
- Semantics: **positive** examples satisfy all clauses; **negative** examples violate one or more clauses
  - 0101 violates $b \rightarrow c$
  - 0111 satisfies all clauses
- **Example**: if target concept is $\{a \rightarrow b, b \rightarrow c, ac \rightarrow d\}$



Preliminaries, II

Representing Horn CNFs

- New representation using "parallel clauses", or **para-clauses**:
  - Add to the **rhs** all propositions implied by antecedent

i	*clauses*	*para-clauses*
1	$a \to b$	$a \to abcd$
2	$b \to c$	$b \to bc$
3	$ac \to d$	~~$ac \to abcd$~~

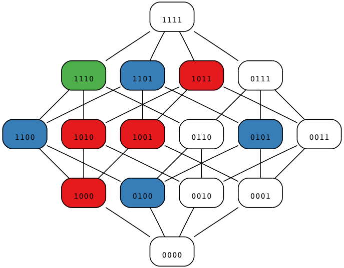
- In new representation:
  - Target concept is $\{a \to abcd, b \to bc\}$, has size 2
  - Para-clause $b \to bc$ meant as $b \to b \wedge b \to c$, so still Horn
  - Omit variables in antecedent: target written as $\{a \to bcd, b \to c\}$
- We prefer new representation because **AFP uses it**!
  - But also: more compact, efficient, ...

Preliminaries, III

Representing Horn CNFs

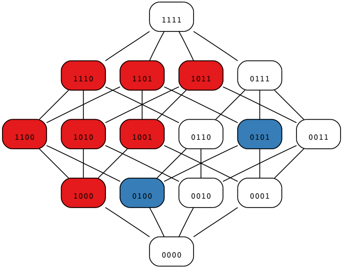
using **clauses**:

$\{a \to b, b \to c, ac \to d\}$



using **para-clauses**:

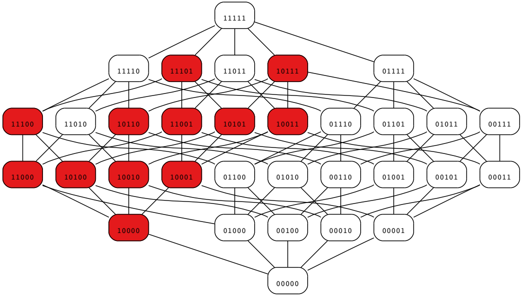
$\{a \to bcd, b \to c\}$



Preliminaries, IV

Lattices

- We learn **finite lower-sub-semi-lattices** $\mathcal{L}$:
  - a **partial order** relation $\leqslant$ on a finite carrier set,
  - the **meet** $\wedge$ (or: greatest lower bound) of two elements is always well-defined



Preliminaries, V

Representing lattices

- Implications in lattices, representable by **comparable pairs**
  - Let x, x' be elements of the lattice $\mathcal{L}$, with $x \leqslant x'$: the elements y that *respect* the comparable pair, denoted $y \models (x, x')$, are those for which the implication "if $x \leqslant y$ then $x' \leqslant y$" holds.

- Example:
  - $(0100, 0110)$ is the implication $b \rightarrow bc$
  - $1100 \not\models (0100, 0110)$ because $0100 \leqslant 1100$ but $0110 \not\leqslant 1100$
  - $1110 \models (0100, 0110)$ because $0100 \leqslant 1110$ and $0110 \leqslant 1110$
  - $1010 \models (0100, 0110)$ because $1010 \not\leqslant 1110$

- Sublattice basis (**subbasis**): a set of comparable pairs, defines the subset of $\mathcal{L}$ consisting of the elements that respect every pair.

- **Cost** is size of minimum basis (in nr. of comparable pairs)

- **Height** is distance between top and bottom of lattice

Preliminaries, VI

Horn CNFs are a special case of lattices

Proposition

Given a set of comparable pairs (a subbasis), **the set of all elements that respect all of them is closed under meet**; and every sublattice has a subbasis, that is, a set of comparable pairs such that the sublattice is exactly the set of elements that respect all of them.

- **Well known fact**: propositional theory is Horn iff it is closed under intersections
- Thus, Horn CNFs are a special case of lattices

lattices	*Horn logic*
meet	bit vector intersection
$\leqslant$	bit-wise $\leqslant$
comparable pair	para-clause
height	nr. of variables n

- Simplifying assumption: we deal with **definite Horn CNFs** or, more in general, sublattices that always contain the top element.

Preliminaries, VII

Closure operator

- Given $x \in \mathcal{L}$, define $x^\star = \bigwedge\{y \in \mathcal{L}^\star \mid x \leqslant y\}$
  - since the target is a sublattice, thus closed under meet, always $x^\star \in \mathcal{L}^\star$, and $x \leqslant x^\star$, with $x = x^\star$ if and only if $x \in \mathcal{L}^\star$
- The $\star$ operator is a **closure** operator:
  - extensive (that is, $x \leqslant x^\star$),
  - monotonic (if $x \leqslant y$ then $x^\star \leqslant y^\star$) and
  - idempotent ($x^{\star\star} = x^\star$); that is, a closure operator.
- What does the closure operator in Horn logic mean?
  - Assume sublattice is given by $\{a \to bcd, b \to c\}$, then
    - $a^\star = abcd$, $b^\star = bc$, $ac^\star = abcd$, $d^\star = d$
    - $1000^\star = 1111$, $0100^\star = 0110$, $1010^\star = 1111$, $0001^\star = 0001$

Preliminaries, VIII

Learning Via Queries

Active Learning

We (the learning algorithm) “interact” with a (hidden) target concept by querying whether a given model is in it and by “showing up for an exam” when we believe we know a correct set of axioms.

More precisely

Identify a target concept through:

- membership queries, give **label** of examples chosen by learning algorithm;
- equivalence queries, tell whether current hypothesis is correct; otherwise, give **counterexample**.

Outline

- Introduction
- AFP-L
- Certificates
- Conclusion

The AFP-L Algorithm, I

Basic observations

- Assume we are learning para-clause $b \to c$
- A **negative** counterexample y gives information about **antecedent** of a para-clause
  - If 1100 is negative, then 1100 satisfies antecedent but falsifies consequent
  - Our first conjecture is $ab \to cd$
    $(1100, 1111)$
- Another **negative** counterexample 0101 gives more information about **antecedent** of a para-clause
  - We check that $1100 \wedge 0101 = 0100$ is still negative (MQ!), if so:
  - Our next conjecture is $b \to acd$
    $(1100 \wedge 0101, 1111) = (0100, 1111)$

Observation

Antecedents are found by intersecting negative examples.

The AFP-L Algorithm, II

Basic observations

- Assume we are learning para-clause $b \to c$
- A **positive** counter-example y tells us about the **consequent** of a para-clause whose antecedent is satisfied by y
  - If 0111 is positive, then rhs of $b \to acd$ must be refined:
  - Our next conjecture is $b \to cd$
    $(0100, 1111 \wedge 0111) = (0100, 0111)$
- Another **positive** counterexample 0110
  - Our next conjecture is $b \to c$
    $(0100, 1111 \wedge 0111 \wedge 0110) = (0100, 0110)$

Observation

Consequents are found by intersecting positive examples.

The AFP-L Algorithm, III

Pseudocode

```
N = empty list
P = {⊤}
t = 0
while EQ(H(N, P)) = ("no", y):
    if y ⊭ H(N, P):
        add y to P
    else: /* N = (x_1, ..., x_t) */
        use MQ to find the first i such that:
            x_i ∧ y is negative
            x_i ∧ y < x_i, that is, x_i ≰ y
        if found, refine: replace x_i by x_i ∧ y in N
        if not found: /* append y to the end of N */
            t = t + 1
            x_t = y
```

Where:

- $P_x = \{y \in P \mid x \leqslant y\}$
- $H(N, P) = \left\{(x, \bigwedge P_x) \mid x \in N\right\}$

The AFP-L Algorithm, IV

A Running Example

- Suppose target concept is: $\{a \rightarrow bcd, b \rightarrow c\}$

	N	P	H(N, P)	EQ?	MQs issued
1	$()$	$\{1111\}$	$\{\}$	1100, $-$	
2	(1100)	$\{1111\}$	$\{ab \rightarrow cd\}$	0100, $-$	$1100 \wedge 0100$, $-$
3	(0100)	$\{1111\}$	$\{b \rightarrow acd\}$	0111, $+$	
4	(0100)	$\{1111, 0111\}$	$\{b \rightarrow cd\}$	1010, $-$	$0100 \wedge 1010$, $+$
5	$(0100, 1010)$	$\{1111, 0111\}$	$\{b \rightarrow cd, ac \rightarrow bd\}$	1001, $-$	$0100 \wedge 1001$, $+$ $1010 \wedge 1001$, $-$
6	$(0100, 1000)$	$\{1111, 0111\}$	$\{b \rightarrow cd, a \rightarrow bcd\}$	0110, $+$	
7	$(0100, 1000)$	$\{1111, 0111, 0110\}$	$\{b \rightarrow c, a \rightarrow bcd\}$	Yes	

The AFP-L Algorithm, V

The new proof

Theorem 1

Let $\mathcal{L}$ be a lattice of height h, and let $\mathcal{L}^\star$ be any sublattice of $\mathcal{L}$ including the top of $\mathcal{L}$, of cost m. Then the algorithm AFP-L learns $\mathcal{L}^\star$ in polynomial time, with at most $O(h \times m^2)$ queries.

To prove theorem we need to demonstrate **termination** in the number of steps stated.

- Algorithm only stops upon receiving a positive answer to an equivalence query

The AFP-L Algorithm, V

The new proof

- $P_x = \{y \in P \mid x \leqslant y\}$

Lemma 1

Let P be a set of positive examples, and let x be a negative example. Consider the comparable pair $(x, \bigwedge P_x)$, and assume that $y \models (x, \bigwedge P_x)$ and that $x \leqslant y$. Then $x^\star \leqslant y$.

- $x^\star \leqslant \bigwedge P_x \leqslant y$ because $P \subseteq \mathcal{L}^\star$, $y \models (x, \bigwedge P_x)$ and $x \leqslant y$ □

Lemma 2 (Invariant)

Along the running of AFP-L, at the point of issuing the equivalence query, for every x_i and x_j in N with $i < j$ there exists an element $z \in \mathcal{L}^\star$ such that $x_i \wedge x_j \leqslant z \leqslant x_j$.

- $N = (x_1, x_2, x_3, \ldots, x_t)$
- By induction on nr of iterations of main loop. □

The AFP-L Algorithm, VI

The new proof

Lemma 3

Let B be a subbasis of the target. Consider two different negative examples x_i and x_j. Each of them has a comparable pair in B that they do not respect: but it cannot be the same pair.

- Suppose by way of contradiction $x_i \not\models (x, y)$ and $x_j \not\models (x, y)$
- $x \leqslant x_i$ and $x \leqslant x_j$ implies $x \leqslant x_i \wedge x_j$
- consider z given by lemma 2 (invariant): $x_i \wedge x_j \leqslant z \leqslant x_j$ and z is positive
- z positive and $x \leqslant x_i \wedge x_j \leqslant z$ implies $y \leqslant z$
- ... **but** $z \leqslant x_j$ so $y \leqslant x_j$ and $x_j \models (x, y)$ (**contradiction!**) □

Theorem 1 follows..

From Lemma 3 and observing the fact that if target has cost m, then length of N never exceeds m (and some counting). □

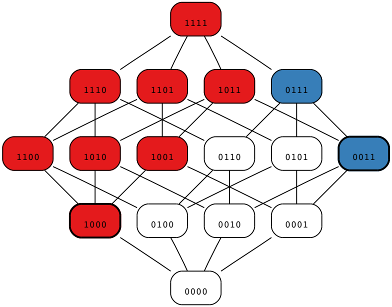
Outline

- Introduction
- AFP-L
- Certificates
- Conclusion

Certificates, I

Definition and Example

- Definition:
  - Let $\mathcal{L}_n^{\leqslant m}$ be a class of sublattices of height n and cost at most m
  - Let f be a concept not in $\mathcal{L}_n^{\leqslant m}$
  - Q_f is a **certificate** for f w.r.t. class $\mathcal{L}_n^{\leqslant m}$ if every sublattice in $\mathcal{L}_n^{\leqslant m}$ differs with f on at least one element in Q_f
- Example: certificates for monotone DNFs, via **minterms**
  - $m = 1, n = 4, f = x_1 + x_3x_4$
  - $Q_f = \{1000+, 0000-, 0011+, 0001-, 0010-\}$
  - Q_f certifies that $f \notin \mathcal{L}_4^{\leqslant 1}$



Certificates, II

Certificates for Lattices

- Given a sublattice $\mathcal{L}^\star \subseteq \mathcal{L}$ of cost m, construct a certificate $Q_{\mathcal{L}^\star}$
- Let $(\alpha_1, \beta_1), \ldots, (\alpha_m, \beta_m)$ be subbasis for $\mathcal{L}^\star$ of minimum cost
- W.l.o.g. assume that $\beta_i = \beta_i^\star$, and thus, $\beta_i = \alpha_i^\star$
- Define $\mathcal{L}_i^\star$ as the sublattice described by the same comparable pairs with the exception of (α_i, β_i).
- Define $x_i = \bigwedge \{y \in \mathcal{L}_i^\star \mid \alpha_i \leqslant y\}$ and $x_i^\star = \bigwedge \{y \in \mathcal{L}^\star \mid \alpha_i \leqslant y\}$

Example

Let $\mathcal{H} = \{e \to d, bc \to d, bd \to c, cd \to b, ad \to bce, ce \to ab\}$

i	$\alpha_i \to \beta_i$	$\alpha_i \to \alpha_i^\star$	x_i	$x_i^\star$
1	$e \to d$	$e \to ed$	00001	00011
2	$bc \to d$	$bc \to bcd$	01100	01110
3	$bd \to c$	$bd \to bcd$	01010	01110
4	$cd \to b$	$cd \to bcd$	00110	01110
5	$ad \to bce$	$ad \to abcde$	10010	11111
6	$ce \to ab$	$ce \to abcde$	01111	11111

Certificates, III

Certificates for Lattices

Lemma 4

For all i, $x_i \not\models (\alpha_i, \beta_i)$; but $x_i \models (\alpha_j, \beta_j)$ for every $j \neq i$.

- Every x_i is negative (by lemma 4)
- Every $x_i^\star$ is positive (by construction)

Lemma 6

If $x_i \leqslant x_j$, then $x_i^\star \leqslant x_j$.

- Because x_j contains all "consequences" of x_i (that is $x_i^\star$) □

Lemma 7

$x_i \wedge x_j \notin \mathcal{L}^\star$ iff $x_i \leqslant x_j$ or $x_j \leqslant x_i$.

- If x_i and x_j incomparable, then $x_i \wedge x_j$ is positive
- If $x_i \leqslant x_j$, then $x_i \wedge x_j = x_i$ is negative □

Certificates, IV

Certificates for Lattices

Theorem

Let $\mathcal{L}^\star$ be a lattice of height n that is not representable by a subbasis of cost m. The set $Q = Q^- \cup Q^+ \cup Q^\wedge$, where

- $Q^- = \{x_i\}_i$, $Q^+ = \{x_i^\star\}_i$, $Q^\wedge = \{x_i \wedge x_j\}_{i<j}$, and $1 \leqslant i, j \leqslant m + 1$

is a certificate set for $\mathcal{L}^\star$of cardinality at most $\binom{m+1}{2} + m + 1$.

- Assume that a sublattice $\mathcal{L}'$ of cost m is consistent with Q
- One comparable pair (α, β) of $\mathcal{L}'$ violated by two different x_i and x_j with $i \neq j$, thus $x_i \wedge x_j \notin \mathcal{L}'$ (pigeonhole principle)
- Since $x_i \not\models (\alpha, \beta)$, we have: $\alpha \leqslant x_i \leqslant x_i^\star$ and $\beta \leqslant x_i^\star$ (otherwise would have $x_i^\star \notin \mathcal{L}'$)
- If x_i and x_j incomparable, then $x_i \wedge x_j \in \mathcal{L}^\star$ (by lemma) but $x_i \wedge x_j \notin \mathcal{L}'$, thus not consistent over $Q^\wedge$ (**contradiction**!)
- If $x_i \leqslant x_j$, then $x_i^\star \leqslant x_j$ (by lemma). But then $\alpha \leqslant x_j$ and $\beta \leqslant x_i^\star \leqslant x_j$ so $x_j \models (\alpha, \beta)$ (**contradiction**!) $\square$

AFP-L and Certificates, I

Simulation of AFP-L using our certificate set

- Let $\mathcal{H} = \{e \to d, bc \to d, bd \to c, cd \to b, ad \to bce, ce \to ab\}$
- In our example simulation, we are going to feed $x_1, x_1^\star, x_2, x_2^\star, x_3, x_4, x_5, x_6$ as counterexamples..

	N	P	EQ(H(N,P))	MQs issued
1	$()$	$\{\top\}$	$x_1 = 00001$	
2	(x_1)	$\{\top\}$	$x_1^\star = 00011$	
3	(x_1)	$\{\top, x_1^\star\}$	$x_2 = 01100$	$x_1 \wedge x_2 = 00000, +$
4	(x_1, x_2)	$\{\top, x_1^\star\}$	$x_2^\star = 01110$	
5	(x_1, x_2)	$\{\top, x_1^\star, x_2^\star\}$	$x_3 = 01010$	$x_1 \wedge x_3 = 00000, +$
				$x_2 \wedge x_3 = 01000, +$
6	(x_1, x_2, x_3)	$\{\top, x_1^\star, x_2^\star\}$	$x_4 = 00110$	$x_1 \wedge x_4 = 00000, +$
				$x_2 \wedge x_4 = 00100, +$
				$x_3 \wedge x_4 = 00010, +$
7	(x_1, x_2, x_3, x_4)	$\{\top, x_1^\star, x_2^\star\}$	$x_5 = 10010$	$x_1 \wedge x_5 = 00000, +$
				$x_2 \wedge x_5 = 00000, +$
				$x_3 \wedge x_5 = 00010, +$
				$x_4 \wedge x_5 = 00010, +$
8	$(x_1, x_2, x_3, x_4, x_5)$	$\{\top, x_1^\star, x_2^\star\}$	$x_6 = 01111$	$x_5 \wedge x_6 = 00010, +$
9	$(x_1, x_2, x_3, x_4, x_5, x_6)$	$\{\top, x_1^\star, x_2^\star\}$	Yes	

AFP-L and Certificates, II

Lemma and Certificates

Lemma 2 (Invariant)

Along the running of AFP-L, at the point of issuing the equivalence query, for every x_i and x_j in N with $i < j$ there exists an element $z \in \mathcal{L}^\star$ such that $x_i \wedge x_j \leqslant z \leqslant x_j$.

- If $x_i \leqslant x_j$, then $z = x_i^\star$
  - $x_i^\star \leqslant x_j$ (by lemma) thus $x_i \wedge x_j = x_i \leqslant x_i^\star \leqslant x_j$
- If x_i and x_j are incomparable, then $z = x_i \wedge x_j$
  - $x_i \wedge x_j \in \mathcal{L}^\star$ (by lemma)

Outline

- Introduction
- AFP-L
- Certificates
- Conclusion

To conclude

In this talk

- The AFP-L learning algorithm
  - Generalization of AFP to lattice setting
  - New, simpler proof
- Certificates for lattices
  - Generalized construction of Horn CNFs to lattice setting
- Better understanding of AFP algorithm
  - Clear link between algorithm, the examples it constructs and certificates

Future work

- How can we use this to make AFP faster?

Thank You! Questions?