Distributive lattice-structured ontologies

Hans Bruun [1] Dion Coumans [2] Mai Gerhke [2]

[1]Technical University of Denmark, Denmark

[2]Radboud University Nijmegen, The Netherlands

CALCO, september 2009

Introduction

Methods of knowledge representation:

- Traditional databases: information stored in tuples

  Name(Dion) - Nat(Dutch) - Prof(PhD-student)

- Ontologies: classification of concepts

  dogs $\subseteq$ animals

Introduction

Methods of knowledge representation:

- Traditional databases: information stored in tuples

  Name(Dion) - Nat(Dutch) - Prof(PhD-student)

- Ontologies: classification of concepts

  dogs $\subseteq$ animals

Aim: provide a common framework (Fischer Nilsson).

Mathematical setting: distributive lattices with additional unary operations preserving $\wedge$, $\vee$ and $\bot$.

Ontological Framework

An Ontological Framework (OF), $\mathcal{O} = (C, A, \Pi)$, consists of:

C set of generators (concept names)

A set of unary operation names (attributes)

Π set of terminological axioms

$$\Pi \subseteq T_{L_\perp A}(C) \times T_{L_\perp A}(C) \qquad (L_\perp A = \{\wedge, \vee, \perp, \{a\}_{a \in A}\})$$

relations on the generators

Problem description

Given $\mathcal{O} = (C, A, \Pi)$, find $\mathbb{L} = (L, \vee, \wedge, \bot, \top, \{a\}_{a \in A})$ such that:

- $(L, \vee, \wedge, \bot, \top)$ is a distributive lattice
- $\mathbb{L}$ is generated by C
- for all $a \in A$,

$$
\begin{array}{lcl}
a(\bot) & = & \bot \\
a(x \vee y) & = & a(x) \vee a(y) \\
a(x \wedge y) & = & a(x) \wedge a(y) \\
a(\top) & = & \top
\end{array}
$$

- $\mathbb{L}$ satisfies the terminological axioms.

Problem description

Given $\mathcal{O} = (C, A, \Pi)$, find $\mathbb{L} = (L, \vee, \wedge, \bot, \top, \{a\}_{a \in A})$ such that:

- $(L, \vee, \wedge, \bot, \top)$ is a distributive lattice
- $\mathbb{L}$ is generated by C
- for all $a \in A$,

$$
\begin{array}{lcl}
a(\bot) & = & \bot \\
a(x \vee y) & = & a(x) \vee a(y) \\
a(x \wedge y) & = & a(x) \wedge a(y) \\
a(\top) & = & \top
\end{array}
$$

- $\mathbb{L}$ satisfies the terminological axioms.

Quotients of $F_{DLA}(C)$ $(= F_{DL}(A^*(C)))$

Solution of an ontological framework

A solution of $\mathcal{O} = (C, A, \Pi)$ is a quotient $h : F_{DLA}(C) \to D$ s.t.
$h(r) = h(s)$ for all $(r, s) \in \Pi$.

Solution of an ontological framework

A solution of $\mathcal{O} = (C, A, \Pi)$ is a quotient $h : F_{DLA}(C) \to D$ s.t.

$h(r) = h(s)$ for all $(r, s) \in \Pi$.

For every ontological framework $\mathcal{O} = (C, A, \Pi)$ there exists a quotient $h_{\mathcal{O}} : F_{DLA}(C) \to F_{\mathcal{O}}$ (universal solution) st:

1. $h_{\mathcal{O}} : F_{DLA}(C) \to F_{\mathcal{O}}$ is a solution of $\mathcal{O} = (C, A, \Pi)$
2. if $h : F_{DLA}(C) \to D$ is any solution of $\mathcal{O} = (C, A, \Pi)$, then there is a unique homomorphism $h_D : F_{\mathcal{O}} \to D$ so that

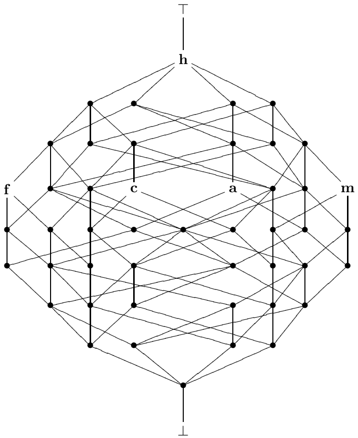
$$\begin{array}{ccc} F_{DLA}(C) & \longrightarrow & D \\ & \searrow \quad \nearrow & \\ & F_{\mathcal{O}} & \end{array}$$

commutes.

Note: The solutions of $\mathcal{O}$ are exactly the quotients of $F_{\mathcal{O}}$.

Example

$$C = \{h, a, c, m, f\}, \quad A = \emptyset, \quad \Pi = \{(h, a \vee c), (h, m \vee f)\}$$



Solutions of an ontological framework

Solutions of an ontological framework $\mathcal{O} = (C, A, \Pi)$:

$h_{\mathcal{O}}:$	$F_{DLA}(C) \to F_{\mathcal{O}}$	universal solution of $\mathcal{O}$	huge
$!:$	$F_{DLA}(C) \to 1$	most collapsed solution	uninteresting

Question: What is a useful solution?

Knowledge Base

A Knowledge Base (KB), $\mathcal{B} = (C, A, \Pi, I)$, consists of

- an ontological framework (C, A, Π)
- a set I of inserted or inhabited terms

  $I \subseteq F_{DLA}(C)$

  terms of interest

Aim: Find the smallest solution of (C, A, Π) in which all information relevant to the inserted terms is present.

Classification of a term

Starting with

$\mathcal{O}$	an ontological framework
$h : F_{DLA}(C) \to D$	a solution of $\mathcal{O}$
$t \in F_{DLA}(C)$	a term

We say $t_1, \ldots, t_n \in F_{DLA}(C)$ is a classification of t w.r.t. h iff

- $t_i \leq t$ for all i
- $h(t) = h(t_1) \vee \ldots \vee h(t_n)$

Classification of a term

We say $t_1, \ldots, t_n \in F_{DLA}(C)$ is a classification of t w.r.t. h iff

- $t_i \leq t \quad$ for all i
- $h(t) = h(t_1) \vee \ldots \vee h(t_n)$

Observe:

- A classification of t wrt the universal solution $F_{\mathcal{O}}$ is also a classification of t wrt any other solution.
- For the trivial quotient, $F_{DLA}(C) \to \mathbf{1}$, $\perp$ is by itself a classification of any term.

Solution of a knowledge base

A solution of $\mathcal{B} = (C, A, \Pi, I)$ is a quotient $h : F_{DLA}(C) \to D$ s.t.

- $h : F_{DLA}(C) \to D$ is a solution of (C, A, Π)
- for all $t \in I$,
  every classification of t wrt h is a classification of t wrt $h_{\mathcal{O}}$

Solution of a knowledge base

A solution of $\mathcal{B} = (C, A, \Pi, I)$ is a quotient $h : F_{DLA}(C) \to D$ s.t.

- $h : F_{DLA}(C) \to D$ is a solution of (C, A, Π)
- for all $t \in I$,
  every classification of t wrt h is a classification of t wrt $h_{\mathcal{O}}$

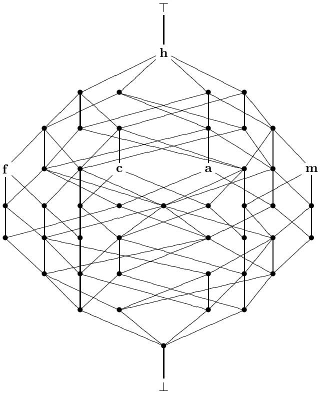
Note: The universal solution (C, A, Π) is a solution of $\mathcal{B}$.

Theorem: Every KB has a least solution $h_{\mathcal{B}} : F_{DLA}(C) \to D_{\mathcal{B}}$:

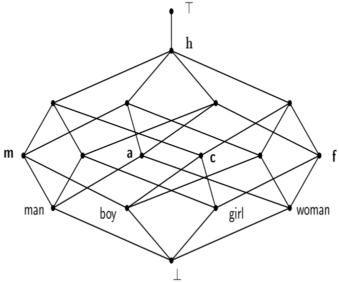
$$\begin{array}{ccc} F_{DLA}(C) & \longrightarrow & D_{\mathcal{B}} \\ & \searrow \quad \nearrow & \\ & D & \end{array}$$

Example

$$C = \{h, a, c, m, f\}, \quad \Pi = \{(h, a \vee c), (h, m \vee f)\}, \quad I = \{h\}$$



(a) Universal solution



(b) Terminal solution

Duality for DLAs

Distributive lattices with add'al operations

$(D, \{a\}_{a \in A})$

Ordered topological spaces with add'al functions

$(P_D, \tau, \leq, \{f_a\}_{a \in A})$

$$P_D = \text{ prime filters of } D$$

$$\tau = \text{ gen. by } \{\hat{d}, (\hat{d})^c : d \in D\}$$

$$\hat{d} = \{\wp \in P_D : d \in \wp\}$$

Duality for DLAs

Distributive lattices with add'al operations

$$(D, \{a\}_{a \in A})$$

Ordered topological spaces with add'al functions

$$(P_D, \tau, \leq, \{f_a\}_{a \in A})$$

P_D = prime filters of D

τ = gen. by $\{\hat{d}, (\hat{d})^c : d \in D\}$

$$\hat{d} = \{\wp \in P_D : d \in \wp\}$$

DLA-quotient

$$D \twoheadrightarrow E$$

Topological closed subset closed under the maps f_a

$$P_E \hookrightarrow P_D$$

Solution of a knowledge base

$h : F_{DLA}(C) \to D$ is a solution of $\mathcal{B} = (C, A, \Pi, I)$ iff

- $h : F_{DLA}(C) \to D$ is a solution of (C, A, Π)

$F_{DLA}(C) \twoheadrightarrow F_{\mathcal{O}} \twoheadrightarrow D$

Quotient of $F_{\mathcal{O}}$

$P_D \hookrightarrow P_{\mathcal{O}} \hookrightarrow P$

Topologically closed subspace of $P_{\mathcal{O}}$ closed under f_a's

- for all $t \in I$, every clas. of t wrt h is a clas. of t wrt $h_{\mathcal{O}}$

for all $t \in I, s \in F_{DLA}(C)$,

$$h(t) \leq h(s) \Rightarrow h_{\mathcal{O}}(t) \leq h_{\mathcal{O}}(s)$$

for all $t \in I$,

$$max(\widehat{h_{\mathcal{O}}(t)}) \subseteq P_D$$

Terminal solution of a knowledge base

For every knowledge base $\mathcal{B} = (C, A, \Pi, I)$ there exists a quotient $h_{\mathcal{B}} : F_{DLA}(C) \to D_{\mathcal{B}}$ (terminal solution) s.t.

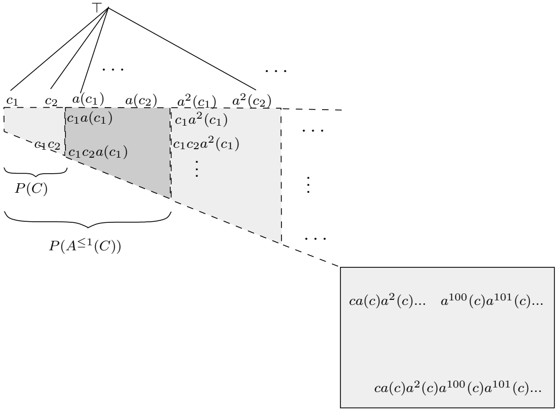
- $h_{\mathcal{B}} : F_{DLA}(C) \to D_{\mathcal{B}}$ is a solution of $\mathcal{B}$
- if $h : F_{DLA}(C) \to D$ is a solution of $\mathcal{B}$, then there exists a unique homomorphism $D \to D_{\mathcal{B}}$ making

$$\begin{array}{ccc} F_{DLA}(C) & \longrightarrow & D_{\mathcal{B}} \\ & \searrow \quad \nearrow & \\ & D & \end{array} \qquad \text{commute}$$

$D_{\mathcal{B}}$ is described dually by

$$P_{\mathcal{B}} = \overline{\left(\bigcup \{ f_w(max(\widehat{h_{\mathcal{O}}(t)})) \mid w \in A^*, \, t \in I \} \right)}$$

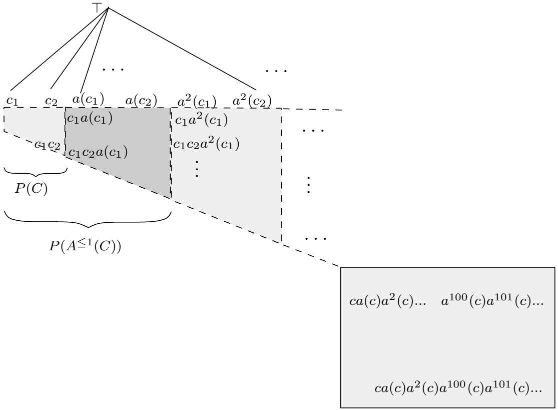
Dual frame of $F_{DLA}(C)$



$$F_{DLA}(C) = F_{DL}(A^*(C))$$

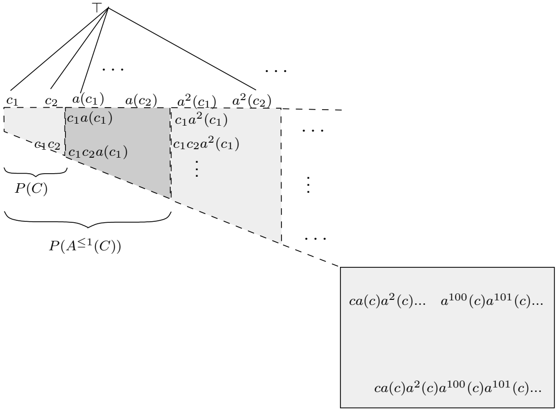
prime filters of $F_{DLA}(C)$ $\leftrightarrow$ subsets of (conjunctions over) $A^*(C)$

Dual frame of $F_{DLA}(C)$



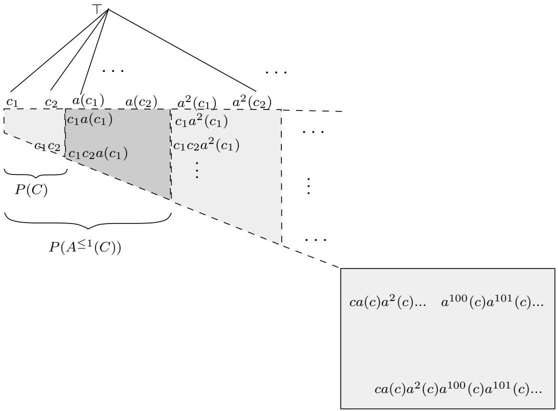
$$P_{\mathcal{O}} = \{p \mid \forall (r,s) \in \Pi \qquad [p \leq \quad r \;\Leftrightarrow p \leq \quad s\;]\}$$

Dual frame of $F_{DLA}(C)$



$$P_{\mathcal{O}} = \{p \mid \forall (r,s) \in \Pi \, \forall w \in A^* \, [\, p \leq w(r) \Leftrightarrow p \leq w(s) \,]\}$$

Dual frame of $F_{DLA}(C)$



$$P_{\mathcal{O}} = \{p \mid \forall (r,s) \in \Pi \, \forall w \in A^* \, [\, p \leq w(r) \Leftrightarrow p \leq w(s) \,]\}$$

$$\widehat{h_{\mathcal{O}}(t)} = \widehat{t} \cap P_{\mathcal{O}}$$

Computing $max(h_{\mathcal{O}}(t))$

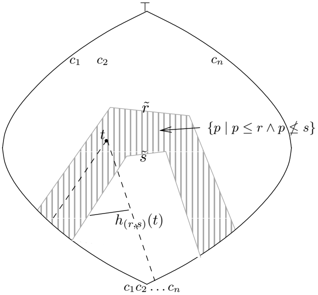
Attribute-free setting

Let $(r, s) \in T^2_{L_\perp A}, \quad r \geq s$

t finite conjunction over $A^*(C)$

$\widehat{h_{(r,s)}(t)} = \widehat{t} \cap P_{(r,s)}$

$P_{(r,s)} = \{p \mid p \leq r \Rightarrow p \leq s\}$



$$max(\widehat{h_{(r,s)})(t)} = \begin{cases} t & \text{if } t \not\leq r \\ max(\{t \wedge p' \mid p' \in \tilde{s}\}) & \text{if } t \leq r \end{cases}$$

To find $max(\widehat{h_{\mathcal{O}}(t)})$ repeatedly apply this basic step.

Computing $max(h_{\mathcal{O}}(t))$

Problem: in the setting with attributes you might not reach $max(\widehat{h_{\mathcal{O}}(t)})$ in finitely many steps.

Example

$$C = \{c\}, \quad A = \{a\}, \quad \Pi = \{(c, c \wedge a(c))\}, \quad I = \{c\}$$

$$c$$

$$\Big\downarrow \; (c, c \wedge a(c))$$

$$c \wedge a(c)$$

$$\Big\downarrow \; (a(c), a(c) \wedge a^2(c))$$

$$c \wedge a(c) \wedge a^2(c)$$

$$\Big\downarrow \; (a^2(c), a^2(c) \wedge a^3(c))$$

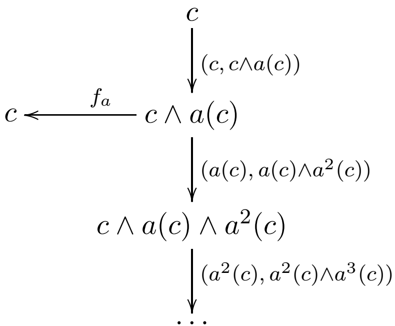
$$\ldots$$

Computing $max(h_{\mathcal{O}}(t))$

Problem: in the setting with attributes you might not reach $max(\widehat{h_{\mathcal{O}}(t)})$ in finitely many steps.

Example

$$C = \{c\}, \quad A = \{a\}, \quad \Pi = \{(c, c \wedge a(c))\}, \quad I = \{c\}$$



Open problems

Using the notion of a classification of a term we showed that every knowledge base has a smallest meaningful solution.

Open problems:

- finding a sharp termination condition
- extension to non-deterministic modalities
- understanding the relation to description logic

Complexity

First problem: a sharp termination condition.

But, under the assumption that $P_{\mathcal{B}}$ is finite (e.g. in attribute-free setting):

- Worst case: size of $P_{\mathcal{B}}$ is exponential.
- In practice the algorithm works well (Oles, Ontoquery project).

Relation to Description Logic

We address a different problem:

- Description Logic: given two terms s, t, decide whether $h_{\mathcal{O}}(s) \leq h_{\mathcal{O}}(t)$.
- Our work: give a decomposition of each relevant term in terms of join irreducibles and thus give a global solution. This in particular generates the relevant join-irreducibles.

Example

$C = \{h, a, c, m, f\}, \quad \Pi = \{(h, a \vee c), (h, m \vee f)\}, \quad I = \{h\}.$