Conservativity of Boolean algebras with operators over semilattices with operators

A. Kurucz, Y. Tanaka*, F. Wolter and M. Zakharyaschev

*Kyushu Sangyo University

TACL 2011

Table of contents

Description logic $\mathcal{EL}$

In this talk, we develop an algebraic semantics for $\mathcal{EL}$.

- $\mathcal{EL}$ is a tractable description logic, and is used for representing large scale ontologies in medicine and other life sciences.
- The profile *OWL 2 EL* of *OWL 2* Web Ontology Language is based on $\mathcal{EL}$.

Example: SNOMED CT – Comprehensive health care terminology with approximately 400,000 definitions.

Examples of concept inclusions of $\mathcal{EL}$:

- $\text{Pericardium} \sqsubseteq \text{Tissue} \sqcap \exists \text{contained\_in}.\text{Heart}$
- $\text{Pericarditis} \sqsubseteq \text{Inflammation} \sqcap \exists \text{has\_location}.\text{Pericardium}$
- $\text{Inflammation} \sqsubseteq \text{Disease} \sqcap \exists \text{acts\_on}.\text{Tissue}$

Concept and Theory of $\mathcal{EL}$

Concepts of $\mathcal{EL}$:

- Two disjoint countably infinite sets NC of *concept names* and NR of *role names*.
- *$\mathcal{EL}$-concepts* C are defined inductively as follows:

$$C ::= \top \mid \bot \mid A \mid C_1 \sqcap C_2 \mid \exists r.C,$$

  where $A \in \mathrm{NC}$, $r \in \mathrm{NR}$ and C_1, C_2 and C are $\mathcal{EL}$-concepts.

Concept inclusions and theories of $\mathcal{EL}$:

- A *concept inclusion* is an expression $C \sqsubseteq D$, where C and D are $\mathcal{EL}$-concepts.
- An *$\mathcal{EL}$-theory* is a set of $\mathcal{EL}$ concept inclusions.

$\sharp$ $\mathcal{EL}$ can be regarded as a fragment of modal logic constructed from propositional variables, $\top$, $\bot$, $\wedge$ and $\Diamond_r$ for each $r \in \mathrm{NR}$.

Interpretation of $\mathcal{EL}$

An *interpretation* of $\mathcal{EL}$ is a structure $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where

- $\Delta^{\mathcal{I}} \neq \varnothing$ is the *domain* of interpretation and
- $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ for each $A \in \mathrm{NC}$ and $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ for each $r \in \mathrm{NR}$.
- $\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$, $\bot^{\mathcal{I}} = \varnothing$.
- $(C_1 \sqcap C_2)^{\mathcal{I}} = C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$.
- $(\exists r.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \exists y \in C^{\mathcal{I}}((x, y) \in r^{\mathcal{I}})\}$.

We say that $\mathcal{I}$ satisfies $C \sqsubseteq D$ and write $\mathcal{I} \models C \sqsubseteq D$, if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

Certain constraints could be put on binary relations $r^{\mathcal{I}}$. Standard constraints on *OWL 2 EL* are transitivity and reflexivity as well as symmetry and functionality.

$\sharp$ Interpretation of $\mathcal{EL}$ can be regarded as a Kripke model, equivalently, a model on a complex Boolean algebra with operators.

Model of $\mathcal{EL}$-theories and quasi-equations

Let $\mathcal{X}$ be an $\mathcal{EL}$-theory. An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *is a model of* $\mathcal{X}$ if it satisfies $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for every $C \sqsubseteq D \in \mathcal{X}$.

Theorem

(Sofronie-Stokkermans 08). For any finite $\mathcal{EL}$-theory $\mathcal{X}$ and any concept inclusion $C \sqsubseteq D$, the following two conditions are equivalent:

- *$C \sqsubseteq D$ is valid in every models of $\mathcal{X}$.*
- BAO $\models \bigwedge \mathcal{X} \to C \sqsubseteq D$, *where BAO is the class of Boolean algebras with operators.*

$\sharp$ Validity of concept inclusions in the models of an $\mathcal{EL}$-theory corresponds to validity of quasi-equations in BAOs.

$\sharp$ What is a proof system, or, in other words, an algebraic semantics for $\mathcal{EL}$?

Algebraic semantics of $\mathcal{EL}$

An algebraic semantics of $\mathcal{EL}$:

- The underlying algebras are bounded meet-semilattices with monotone operators f_r for each $r \in \mathrm{NR}$ (SLOs, for short).
- An $\mathcal{EL}$ concept is interpreted as a term of the language of SLOs.
- A concept inclusion $C \sqsubseteq D$ is interpreted as an equation $C \leq D$.
- Relational constraints of original interpretation are given by equational theories of SLO. For example, $x \leq fx$ for reflexivity.

$\sharp$ Is the SLO semantics equivalent to original interpretation for $\mathcal{EL}$?

Conservativity and completeness

Let $\mathcal{C}$ denotes the class of algebras, $\mathcal{T}$ a set of equations of SLO and $\mathbf{q}$ a quasi-equation of SLO. We say

- $\mathcal{T} \models_{\mathcal{C}} \mathbf{q}$ if $\mathfrak{A} \models \mathbf{q}$ for every $\mathfrak{A} \in \mathcal{C}$ with $\mathfrak{A} \models \mathcal{T}$;
- $\mathcal{T}$ is $\mathcal{C}$-*conservative* if $\mathcal{T} \models_{\mathcal{C}} \mathbf{q}$ implies $\mathcal{T} \models_{\mathsf{SLO}} \mathbf{q}$ for every $\mathbf{q}$;
- $\mathcal{T}$ is *complete* if it is CA-conservative, where CA is the set of all complex Boolean algebras with operators.

Theorem

(Sofronie-Stokkermans 08). Any subset of the following theory is complete:

$$\{f_{r_2} \circ f_{r_1}(x) \leq f_r(x) \mid r_1,\ r_2,\ r \in \mathrm{NR}\} \cup \{f_r(x) \leq f_s(x) \mid r,\ s \in \mathrm{NR}\}$$

♯ Completeness of $\{ffx \leq fx\}$ for transitivity follows from the above theorem.

♯ Which relational constraints are complete?

Completeness and embedding

We give relational constraints of original interpretation by equational theories $\mathcal{T}$ of SLO. Is it complete with respect to the original interpretation?

Let $\mathsf{V}(\mathcal{T})$ be the variety of SLOs axiomatized by $\mathcal{T}$. We say that $\mathcal{T}$ is *complex* if every $\mathfrak{A} \in \mathsf{V}(\mathcal{T})$ is embeddable in a complex BAO $\mathfrak{B}$ whose reduct to SLO is in $\mathsf{V}(\mathcal{T})$.

Theorem

For every $\mathcal{T}$, the following conditions are equivalent:

1. *$\mathcal{T}$ is complex.*
2. *$\mathcal{T}$ is complete. ($\mathcal{T} \models_{\mathsf{CA}} \mathbf{q} \Rightarrow \mathcal{T} \models_{\mathsf{SLO}} \mathbf{q}$.)*
3. *$\mathcal{T}$ is BAO-conservative. ($\mathcal{T} \models_{\mathsf{BAO}} \mathbf{q} \Rightarrow \mathcal{T} \models_{\mathsf{SLO}} \mathbf{q}$.)*

♯ So, if we find an appropriate embedding, we get completeness.

Constructing embeddings

We construct an embedding via two steps:

1. Embed any SLO validating $\mathcal{T}$ into a DLO validating $\mathcal{T}$:
   This is equivalent to prove DLO-conservativity, that is,
   $$\mathcal{T} \models_{\mathrm{DLO}} \mathbf{q} \Rightarrow \mathcal{T} \models_{\mathrm{SLO}} \mathbf{q}.$$
2. Embed any DLO validating $\mathcal{T}$ into a BAO validating $\mathcal{T}$:
   This is equivalent to prove DLO-BAO-conservativity, that is,
   $$\mathcal{T} \models_{\mathrm{BAO}} \mathbf{q} \Rightarrow \mathcal{T} \models_{\mathrm{DLO}} \mathbf{q}.$$

Embedding SLO into DLO

As concerns for embedding from SLOs into DLOs, we have the following result:

Theorem

Every $\mathcal{EL}$-theory containing only equations where each variable occurs at most once in the left-hand side is DLO-conservative.

Example: An $\mathcal{EL}$-theory $\mathcal{T}_{\mathcal{S}5}$ satisfies the condition of the theorem, but $\mathcal{T}_{\mathcal{S}4.3}$ does not, where

$$\mathcal{T}_{\mathcal{S}5} = \{x \leq fx,\ ffx \leq fx,\ x \wedge fy \leq f(fx \wedge y)\}$$

$$\mathcal{T}_{\mathcal{S}4.3} = \{x \leq fx,\ ffx \leq fx,\ f(x \wedge y) \wedge f(x \wedge z) \leq f(x \wedge fy \wedge fz)\}.$$

$\sharp$ As we will see later, $\mathcal{T}_{\mathcal{S}4.3}$ is not DLO-conservative.

Embedding DLO into BAO

Embedding from a DLO $\mathfrak{D}$ to a BAO is given by defining appropriate binary relation R on the set $\mathcal{F}(\mathfrak{D})$ of prime filters of $\mathfrak{D}$.

Let $\mathfrak{B}$ be the complex BA defined on the set $\wp(\mathcal{F}(\mathfrak{D}))$. Let $f_{\mathfrak{D}}$ be the operator on $\mathfrak{D}$ and $f_{\mathfrak{B}}$ an operator on $\mathfrak{B}$ defined by $f_{\mathfrak{B}}(U) = \{F \mid \exists G \in U \ (F, G) \in R\}$.

Example:

- If $f_{\mathfrak{D}}$ is functional and $(F, G) \in R \Leftrightarrow G = f_{\mathfrak{D}}^{-1}(F)$, then $f_{\mathfrak{B}}$ is functional.
- If $f_{\mathfrak{D}}$ is symmetry and $(F, G) \in R \Leftrightarrow f_{\mathfrak{D}}(G) \subseteq F$ and $f_{\mathfrak{D}}(F) \subseteq G$, then $f_{\mathfrak{B}}$ is symmetry.

$\sharp$ Unfortunately, we don't know any general way to define R.

Complete theories

As a consequence, we have following completeness results:

Theorem

The following $\mathcal{EL}$-theories are complete:

- *Symmetry:*

$$\{x \wedge fy \leq f(fx \wedge y)\}$$

- *Functionality:*

$$\{fx \wedge fy \leq f(x \wedge y)\}$$

- *Reflexivity, transitivity and symmetry:*

$$\mathcal{T}_{\mathcal{S}5} = \{x \leq fx,\ ffx \leq fx,\ x \wedge fy \leq f(fx \wedge y)\}$$

Fusion of $\mathcal{EL}$ theories

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be $\mathcal{EL}$-theories. We call $\mathcal{T}_1 \cup \mathcal{T}_2$ a *fusion* of $\mathcal{T}_1$ and $\mathcal{T}_2$ if the set of f-operators occurring in $\mathcal{T}_1$ and $\mathcal{T}_2$ are disjoint.

Theorem

The fusions of complete $\mathcal{EL}$-theories are also complete.

$\sharp$ Union of complete theories is not complete in general, as we will see later.

Incompleteness

There are $\mathcal{EL}$ theories $\mathcal{T}$ which are incomplete. That is, there exists quasi-equation $\mathbf{q}$ such that

$$\mathcal{T} \models_{\mathsf{CA}} \mathbf{q}, \ \mathcal{T} \not\models_{\mathsf{SLO}} \mathbf{q}.$$

Some incomplete $\mathcal{EL}$ theories are DLO-nonconservative. That is, there exists quasi-equation $\mathbf{q}$ such that

$$\mathcal{T} \models_{\mathsf{DLO}} \mathbf{q}, \ \mathcal{T} \not\models_{\mathsf{SLO}} \mathbf{q}.$$

BAO-nonconservative incomplete $\mathcal{EL}$ theory

Example: Both $\{x \leq fx\}$ and $\{fx \wedge fy \leq f(x \wedge y)\}$ are complete, but their union is not. Let $\mathfrak{S} = \{0, a, 1\}$, $f0 = 0$ and $fa = f1 = 1$. Then, $fa \nleq a$. However, in BAO

$$\{x \leq fx,\ fx \wedge fy \leq f(x \wedge y)\} \vDash_{\mathrm{BAO}} fx \leq x$$

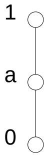

Figure: $fa \nleq a$

♯ On the other hand, the above theory is DLO-conservative.
♯ Union of complete theories is not complete, in general.

DLO-nonconservative incomplete $\mathcal{EL}$ theory

Example: $\mathcal{T}_{\mathcal{S}4.3}$ is DLO-nonconservative and hence incomplete. Let $\mathfrak{S}$ be the following SLO, where $fa = d$, $fc = e$ and $fx = x$ for the remaining x. Then, $a \wedge fc = fa \wedge c$ and $fa \wedge fc \nleq f(a \wedge c)$. However, in DLO

$$\mathcal{T}_{\mathcal{S}4.3} \vDash_{\mathrm{DLO}} x \wedge fy = fx \wedge y \Rightarrow fx \wedge fy \leq f(x \wedge y).$$

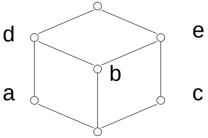

Figure: $a \wedge fc = fa \wedge c$, $fa \wedge fc \nleq f(a \wedge c)$

♯ Is there any SLO equation **e** such that

$$\mathcal{T}_{\mathcal{S}4.3} \vDash_{\mathrm{DLO}} \mathbf{e} \text{ and } \mathcal{T}_{\mathcal{S}4.3} \nvDash_{\mathrm{SLO}} \mathbf{e}?$$

Subvarieties of $\mathcal{S}5$

It is known that the lattice of subvarieties of $\mathrm{V}(\mathcal{T}_{\mathcal{S}5})$ is the following (Jackson 04), where

$$\mathcal{T}_{\mathcal{S}5} = \{x \leq fx,\ ffx \leq fx,\ x \wedge fy \leq f(fx \wedge y)\}.$$

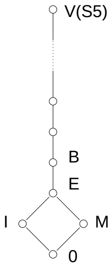

Figure: Lattice of subvarieties of $\mathrm{V}(\mathcal{T}_{\mathcal{S}5})$

Subvarieties of $\mathcal{S}5$

The only incomplete one is $\mathcal{E}$, which is defined by

$$\mathcal{T}_{\mathcal{S}5} \cup \{fx \wedge fy \leq f(x \wedge y)\}.$$

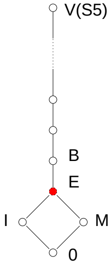

Figure: Lattice of subvarieties of $\mathrm{V}(\mathcal{T}_{\mathcal{S}5})$

Completeness problem for $\mathcal{EL}$-theories

- We have observed that some theories of $\mathcal{EL}$ are complete and some are not.
- So, it is a natural question that whether we can decide a given $\mathcal{EL}$-theory is complete or not.
- The last topic of this presentation is undecidability of this completeness problem for $\mathcal{EL}$-theories.

Undecidability of completeness

By reducing the halting problem for Turing machines, we can show the following:

Theorem

No algorithm can decide, given a finite set $\mathcal{T}$ of $\mathcal{EL}$-equations, whether $\mathcal{T} \vDash_{\mathsf{SLO}} 0 = 1$.

We also have the following:

Theorem

For every $\mathcal{EL}$-theory $\mathcal{T}$, the following two conditions are equivalent:

- *the fusion of $\mathcal{T}$ and $\{f(x) \le x\}$ is complete;*
- $\mathcal{T} \vDash_{\mathsf{SLO}} 0 = 1$.

Undecidability of completeness

Hence, we have undecidability of completeness:

Theorem

It is undecidable whether a finite set $\mathcal{T}$ of $\mathcal{EL}$-equations is complete.

Further research

- General sufficient syntactic criteria for completeness.
- Discuss conservativity for equations, instead of quasi-equations.
- Relation between quasi-varieties of SLOs and varieties of SLOs defined by $\mathcal{EL}$ theories.

Thank you for your attention.