Conservativity of Boolean algebras with operators over semilattices with operators

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

*Kyushu Sangyo University

August 11, 2011
Table of contents

Introduction
  Background and motivation
  Algebraic semantics for $\mathcal{EL}$

 Conservativity and completeness
  Conservativity, completeness and embedding
  Some completeness and incompleteness results
  $\mathcal{EL}$-theories over $S5$

 Undecidability of completeness
  Undecidability of completeness
  Further research
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 $\text{OWL 2 EL}$ of $\text{OWL 2}$ 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 $\mathcal{EL}$:
- Pericardium $\sqsubseteq$ Tissue $\sqcap \exists$ contained_in.Heart
- Pericarditis $\sqsubseteq$ Inflammation $\sqcap \exists$ has_location.Pericardium
- Inflammation $\sqsubseteq$ Disease $\sqcap \exists$ acts_on.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:

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

where $A \in NC$, $r \in NR$ 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.

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

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

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

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

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

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 $C \subseteq D \in \mathcal{X}$.

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

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

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

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 $r \in \text{NR}$ (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.

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 $\mathcal{A} \models \mathbf{q}$ for every $\mathcal{A} \in \mathcal{C}$ with $\mathcal{A} \models \mathcal{T}$;
- $\mathcal{T}$ is $\mathcal{C}$-conservative if $\mathcal{T} \models_{\mathcal{C}} \mathbf{q}$ implies $\mathcal{T} \models_{\text{SLO}} \mathbf{q}$ 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:

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

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

Theorem

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

1. $\mathcal{T}$ is complex.
2. $\mathcal{T}$ is complete. ($\mathcal{T} \vdash_{\text{CA}} q \Rightarrow \mathcal{T} \vdash_{\text{SLO}} q$.)
3. $\mathcal{T}$ is BAO-conservative. ($\mathcal{T} \vdash_{\text{BAO}} q \Rightarrow \mathcal{T} \vdash_{\text{SLO}} q$.)

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,

   $$\mathcal{T} \vdash_{\text{DLO}} q \Rightarrow \mathcal{T} \vdash_{\text{SLO}} q.$$

2. Embed any DLO validating $\mathcal{T}$ into a BAO validating $\mathcal{T}$: This is equivalent to prove DLO-BAO-conservativity, that is,

   $$\mathcal{T} \vdash_{\text{BAO}} q \Rightarrow \mathcal{T} \vdash_{\text{DLO}} q.$$
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 \( \mathcal{T}_{S5} \) satisfies the condition of the theorem, but \( \mathcal{T}_{S4.3} \) does not, where

\[
\mathcal{T}_{S5} = \{ x \leq fx, \ ff x \leq fx, \ x \land fy \leq f(fx \land y) \}
\]

\[
\mathcal{T}_{S4.3} = \{ x \leq fx, \ ff x \leq fx, \ f(x \land y) \land f(x \land z) \leq f(x \land fy \land fz) \}.
\]

As we will see later, \( \mathcal{T}_{S4.3} \) is not DLO-conservative.
Embedding DLO into BAO

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

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

Example:

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

Unfortunately, we don’t know any general way to define $R$. 
Complete theories

As a consequence, we have following completeness results:

Theorem
The following $\mathcal{E}\mathcal{L}$-theories are complete:

- **Symmetry:**
  \[ \{ x \land f y \leq f ( f x \land y ) \} \]

- **Functionality:**
  \[ \{ f x \land f y \leq f ( x \land y ) \} \]

- **Reflexivity, transitivity and symmetry:**
  \[ T_{S5} = \{ x \leq f x, f f x \leq f x, x \land f y \leq f ( f x \land y ) \} \]
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.*

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 $q$ such that

$$\mathcal{T} \vDash_{CA} q, \; \mathcal{T} \nvDash_{SLO} q.$$ 

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

$$\mathcal{T} \vDash_{DLO} q, \; \mathcal{T} \nvDash_{SLO} q.$$
Example: Both $\{x \leq fx\}$ and $\{fx \land fy \leq f(x \land y)\}$ are complete, but their union is not. Let $\mathcal{G} = \{0, a, 1\}$, $f0 = 0$ and $fa = f1 = 1$. Then, $fa \notin a$. However, in BAO

$$\{x \leq fx, \; fx \land fy \leq f(x \land y)\} \models_{\text{BAO}} fx \leq x$$

Figure: $fa \notin a$

On the other hand, the above theory is DLO-conservative.

Union of complete theories is not complete, in general.
DLO-nonconservative incomplete $\mathcal{E}\mathcal{L}$ theory

**Example**: $\mathcal{T}_{S4.3}$ is DLO-nonconservative and hence incomplete. Let $\mathcal{G}$ be the following SLO, where $fa = d$, $fc = e$ and $fx = x$ for the remaining $x$. Then, $a \land fc = fa \land c$ and $fa \land fc \not\leq f(a \land c)$. However, in DLO

$$\mathcal{T}_{S4.3} \vdash_{DLO} x \land fy = fx \land y \Rightarrow fx \land fy \not\leq f(x \land y).$$

![Diagram](image)

**Figure**: $a \land fc = fa \land c$, $fa \land fc \not\leq f(a \land c)$

Is there any SLO equation $e$ such that

$$\mathcal{T}_{S4.3} \vdash_{DLO} e \text{ and } \mathcal{T}_{S4.3} \not\vdash_{SLO} e?$$
Subvarieties of $S_5$

It is known that the lattice of subvarieties of $V(T_{S_5})$ is the following (Jackson 04), where

$$T_{S_5} = \{ x \leq fx, \quad ffx \leq fx, \quad x \land fy \leq f(fx \land y) \}. $$

![Lattice of subvarieties of $V(T_{S_5})$](image)

**Figure:** Lattice of subvarieties of $V(T_{S_5})$
Subvarieties of $S_5$

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

$$\mathcal{T}_{S_5} \cup \{fx \land fy \leq f(x \land y)\}.$$
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 $\mathcal{T} \models_{\text{SLO}} 0 = 1$.

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 $\{f(x) \leq x\}$ is complete;
- $\mathcal{T} \models_{\text{SLO}} 0 = 1$. 
Hence, we have undecidability of completeness:

**Theorem**

*It is undecidable whether a finite set \( T \) 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.