Can you tell the difference between *DL-Lite* ontologies?

Roman Kontchakov

SCSIS, Birkbeck College, London

`http://www.dcs.bbk.ac.uk/~roman`

Frank Wolter and Michael Zakharyaschev

University of Liverpool   Birkbeck College, London

Developing and maintaining ontologies

- **versions:**
  comparing **logical consequences** over some common vocabulary Σ
  not not the syntactic form of the axioms (as in `diff`)

- **refinement:**
  adding new axioms but **preserving** the relationships
  between terms of a certain part Σ of the vocabulary

- **reuse:**
  importing an ontology and using its vocabulary Σ as originally defined
  (relationships between terms of Σ should not change)

new types of reasoning problems

related notions: conservative extensions, model conservativity, locality, etc.

Σ-difference

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be TBoxes (in some DL $\mathcal{L}$) and Σ a **signature** (concept and role names)

Σ-**concept difference** $\mathbf{cDiff}_\Sigma^{\mathcal{L}}(\mathcal{T}_1, \mathcal{T}_2)$ is the set of Σ-concept inclusions such that

$$\mathcal{T}_2 \models C \sqsubseteq D \ \text{ and } \ \mathcal{T}_1 \not\models C \sqsubseteq D$$

Σ-**query difference** $\mathbf{qDiff}_\Sigma^{\mathcal{L}}(\mathcal{T}_1, \mathcal{T}_2)$ is the set of pairs $(\mathcal{A}, q(\vec{x}))$, where

$$\mathbf{sig}(\mathcal{A}), \mathbf{sig}(q) \subseteq \Sigma, \qquad (\mathcal{T}_2, \mathcal{A}) \models q(\vec{a}) \ \text{ and } \ (\mathcal{T}_1, \mathcal{A}) \not\models q(\vec{a}), \ \text{ for some } \vec{a}$$

strong Σ-query difference $\mathbf{sqDiff}_\Sigma^{\mathcal{L}}(\mathcal{T}_1, \mathcal{T}_2)$ is the set of triples $(\mathcal{T}, \mathcal{A}, q(\vec{x}))$, where

$$\mathbf{sig}(\mathcal{T}, \mathcal{A}), \mathbf{sig}(q) \subseteq \Sigma, \quad (\mathcal{T}_2 \cup \mathcal{T}, \mathcal{A}) \models q(\vec{a}), \ (\mathcal{T}_1 \cup \mathcal{T}, \mathcal{A}) \not\models q(\vec{a}), \ \text{ for some } \vec{a}$$

modules: replacing an ontology in a **context** $\mathcal{T}$

$\mathcal{T}_1$ and $\mathcal{T}_2$ are $\Sigma\text{-}{}^{\text{concept}}_{\substack{\text{query} \\ \text{strong query}}}$ **inseparable** iff ${}^{\mathsf{c}}_{\substack{\mathsf{q} \\ \mathsf{sq}}}\mathbf{Diff}(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$ and ${}^{\mathsf{c}}_{\substack{\mathsf{q} \\ \mathsf{sq}}}\mathbf{Diff}(\mathcal{T}_2, \mathcal{T}_1) = \emptyset$

— ExpTime for $\mathcal{EL}$, 2ExpTime for $\mathcal{ALCQI}$, undecidable for $\mathcal{ALCQIO}$
$+$ tractable for acyclic $\mathcal{EL}$ (e.g., SNOMED)

DL-Lite: Description Logic for Databases

A fragment of a conceptual schema:

Translating into DL:

$\exists\,\mathsf{manages}.\top \;\sqsubseteq\; \mathsf{ProjectManager}$

$\exists\,\mathsf{manages}^-.\top \;\sqsubseteq\; \mathsf{Project}$

$\mathsf{Project} \;\sqsubseteq\; \exists\,\mathsf{manages}^-.\top$

$\geq 3\ \mathsf{manages}^-.\top \;\sqsubseteq\; \bot$

$\mathsf{Research} \sqcap \mathsf{Visiting} \;\sqsubseteq\; \bot$

$\mathsf{Academic} \;\sqsubseteq\; \mathsf{ProjectManager}$

$\mathsf{ProjectManager} \;\sqsubseteq\; \mathsf{Academic} \sqcup \mathsf{Visiting}$

. . .

DL-Lite$_{horn}$	$B_1 \sqcap \cdots \sqcap B_k \sqsubseteq B$	*DL-Lite*$_{bool}$	$C_1 \sqsubseteq C_2$

$$
\begin{aligned}
B &::= \quad \bot \quad | \quad A_i \quad | \quad \exists R \quad | \quad \geq q\,R \\
C &::= \quad B \quad | \quad \neg C \quad | \quad C_1 \sqcap C_2 \quad | \quad C_1 \sqcup C_2
\end{aligned}
$$

Example

Let $\mathcal{T}_1$ contain the axioms

$$\text{Research} \sqsubseteq \exists\text{worksIn}, \qquad\qquad \exists\text{worksIn}^- \sqsubseteq \text{Project},$$
$$\text{Project} \sqsubseteq \exists\text{manages}^-, \qquad\qquad \exists\text{manages} \sqsubseteq \text{Academic} \sqcup \text{Visiting},$$
$$\exists\text{teaches} \sqsubseteq \text{Academic} \sqcup \text{Research}, \quad \text{Academic} \sqsubseteq \exists\text{teaches} \sqcap \,\leq 1\,\text{teaches},$$
$$\text{Research} \sqcap \text{Visiting} \sqsubseteq \bot, \qquad\qquad \exists\text{writes} \sqsubseteq \text{Academic} \sqcup \text{Research},$$

$\mathcal{T}_2 = \mathcal{T}_1 \cup \{\text{Visiting} \sqsubseteq\, \geq 2\,\text{writes}\}$ and $\Sigma = \{\text{teaches}\}$

- $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-concept inseparable   (Σ-entailment in both directions)
  $\mathcal{T}_2 \models \text{Visiting} \sqsubseteq \text{Academic}$, but nothing new in the signature Σ

- $\mathcal{T}_1$ does not Σ-query entail $\mathcal{T}_2$:
  $\mathcal{A} = \{\text{teaches}(a, b), \text{teaches}(a, c)\}$
  $q = \exists x\,((\exists\text{teaches})(x) \wedge (\leq 1\,\text{teaches})(x))$
  'is there anybody who teaches precisely one module?'

$$(\mathcal{T}_1, \mathcal{A}) \not\models q$$
$$(\mathcal{T}_2, \mathcal{A}) \models q$$

Σ-inseparability in *DL-Lite*

Theorem

(1) In *DL-Lite$_{bool}$*:

Strong Σ-query insep. $\Leftrightarrow$ Σ-query inseparability $\Rightarrow$ Σ-concept inseparability

In each case the problem is Π_2^p-complete

(2) In *DL-Lite$_{horn}$*:

Strong Σ-query insep. $\Rightarrow$ Σ-query inseparability $\Rightarrow$ Σ-concept inseparability

In each case the problem is **coNP**-complete

(3) In *DL-Lite$_{bool}$*:

Σ-query entailment and Σ-concept entailment

can be encoded by Quantified Boolean Formulas $\forall\exists\,\psi$

Σ-entailment: semantic criteria

Let Q be a set of numerical parameters and Σ a signature

ΣQ-concepts B: $A_i \in \Sigma$ and $(\geq q\, R)$ with $q \in Q$ and $R \in \Sigma$

> ΣQ-type $\boldsymbol{t}$ is a set of ΣQ-concepts containing
> - B or $\neg B$ (but not both), for all B
> - $\geq q\, R$ whenever $q < q'$ and $\geq q'\, R \in \boldsymbol{t}$, for all $\geq q\, R$

For a TBox $\mathcal{T}$,

  a ΣQ-type $\boldsymbol{t}$ is **$\mathcal{T}$-realisable** if $\boldsymbol{t}$ is satisfied in a model of $\mathcal{T}$

  a set Ξ of ΣQ-types is **precisely $\mathcal{T}$-realisable** if
            there is a model of $\mathcal{T}$ realising precisely the types from Ξ

Theorem. Let Q denote the set of parameters occurring in $\mathcal{T}_1 \cup \mathcal{T}_2$

  $\mathcal{T}_1$ **Σ-concept entails** $\mathcal{T}_2$ iff   every $\mathcal{T}_1$-realisable ΣQ-type is **$\mathcal{T}_2$-realisable**

  $\mathcal{T}_1$ **Σ-query entails** $\mathcal{T}_2$ iff   every precisely $\mathcal{T}_1$-realisable set Ξ of ΣQ-types
                              is **precisely $\mathcal{T}_2$-realisable**

Encoding Σ-concept entailment in QBF

Let $\mathcal{T}$ be a TBox, Q a set of numerical parameters and $\boldsymbol{t}$ a $\mathbf{sig}(\mathcal{T})Q$-type

'$\boldsymbol{t}_0$ is $\mathcal{T}$-realisable with $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ being witnesses' propositional formula	$=$	$\Phi_{\mathcal{T}}(b_0, b_1, \ldots, b_n)$

b_j is the vector of all propositional variables B^* of the type $\boldsymbol{t}_j$

Then the condition

'every $\mathcal{T}_1$-realisable ΣQ-type $\boldsymbol{t}$ is $\mathcal{T}_2$-realisable'

is described by the following QBF

$$\forall b_0^{\Sigma Q} \Big[\exists b_0^{\mathcal{T}_2 \setminus \Sigma Q} \exists b_1^{\mathcal{T}_1} \ldots \exists b_{n_1}^{\mathcal{T}_1} \ \Phi_{\mathcal{T}_1}(b_0^{\Sigma Q} \cdot b_0^{\mathcal{T}_1 \setminus \Sigma Q}, b_1^{\mathcal{T}_1}, \ldots, b_{n_1}^{\mathcal{T}_1}) \quad \rightarrow$$
$$\exists b_0^{\mathcal{T}_2 \setminus \Sigma Q} \exists b_1^{\mathcal{T}_2} \ldots \exists b_{n_2}^{\mathcal{T}_2} \ \Phi_{\mathcal{T}_2}(b_0^{\Sigma Q} \cdot b_0^{\mathcal{T}_2 \setminus \Sigma Q}, b_1^{\mathcal{T}_2}, \ldots, b_{n_2}^{\mathcal{T}_2}) \Big]$$

($b_0^{\Sigma Q}$ is the ΣQ-part of b_0 and $b_0^{\mathcal{T}_i \setminus \Sigma Q}$ contains the rest of the variables)

Experiments

TBox instances   (standard Department Ontology + ICNARC)

series	description	no. of instances	axioms $\mathcal{T}_1$	axioms $\mathcal{T}_2$	basic concepts $\mathcal{T}_1$	basic concepts $\mathcal{T}_2$	Σ
NN	$\mathcal{T}_1$ does not Σ-concept entail $\mathcal{T}_2$	420	59–154	74–198	47–121	49–146	5–52
YN	$\mathcal{T}_1$ Σ-concept but not Σ-query entails $\mathcal{T}_2$	252	56–151	77–191	44–119	58–145	6–45
YY	$\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$	156	54–88	62–110	43–79	47–94	6–32

QBF solvers

- sKizzo 0.8.2 (http://skizzo.info/).
- 2clsQ (1st place QBF Competition 2006, http://www.cs.toronto.edu/~fbacchus/)
- yQuaffle (http://www.princeton.edu/~chaff/quaffle.html)
- QuBE 6.4 (http://www.star.dist.unige.it/)

series	Σ-concept entailment QBF variables	Σ-concept entailment QBF clauses	Σ-query entailment QBF variables	Σ-query entailment QBF clauses
NN	1,469–11,752	2,391–18,277	1,715–15,174	5,763–163,936
YN	1,460–11,318	2,352–17,424	1,755–14,723	7,006–151,452
YY	1,526–4,146	2,200–6,079	1,510–4,946	5,121–29,120

Experimental results: percentage of solved instances

Σ-concept entailment

Σ-query entailment

NN

YN

YY

2clsQ	sKizzo	yQuaffle	QuBE

Forgetting

studied under different names: forgetting, uniform interpolation, variable elimination. . .

A DL $\mathcal{L}$ **admits forgetting** (**has uniform interpolation**) if,
for every $\mathcal{T}$ in $\mathcal{L}$ and every Σ, there exists $\mathcal{T}_\Sigma$ in $\mathcal{L}$ with **sig**$(\mathcal{T}_\Sigma) \subseteq \Sigma$
such that $\mathcal{T}$ and $\mathcal{T}_\Sigma$ are Σ-concept inseparable in $\mathcal{L}$

Theorem Both *DL-Lite$_{bool}$* and *DL-Lite$_{horn}$* have uniform interpolation
and the uniform interpolant can be constructed in exponential time

DL-Lite$_{bool}^u$:

$$C \quad ::= \quad \ldots \quad | \quad \exists C \quad | \quad \ldots$$

(**universal modality**)

e.g., $(\geq 2 \text{ teaches}) \sqsubseteq \exists(\exists \text{teaches} \sqcap \leq 1 \text{ teaches})$

$\mathcal{T}_\Sigma$ with **sig**$(\mathcal{T}_\Sigma) \subseteq \Sigma$ is a **uniform interpolant** of $\mathcal{T}$ w.r.t. Σ in ***DL-Lite$_{bool}^u$*** if
$\mathcal{T} \models C \sqsubseteq D$ iff $\mathcal{T}_\Sigma \models C \sqsubseteq D$, for every $C \sqsubseteq D$ in *DL-Lite$_{bool}^u$* with **sig**$(C \sqsubseteq D) \sqsubseteq \Sigma$

$\boxed{\mathcal{T}' \ \Sigma\text{-query entails } \mathcal{T} \quad \text{iff} \quad \mathcal{T}' \models C \sqsubseteq D, \text{ for each } C \sqsubseteq D \in \mathcal{T}_\Sigma}$

Theorem For every $\mathcal{T}$ in *DL-Lite$_{bool}$* and every Σ one can construct
a uniform interpolant $\mathcal{T}_\Sigma$ of $\mathcal{T}$ w.r.t. Σ in *DL-Lite$_{bool}^u$* in time exponential in $\mathcal{T}$

Future work

- investigate different variants of the QBF encoding   (non-prenex/non-CNF)
                    and/or different solvers   (AQME or even a dedicated solver)

- QBF encoding of Σ-entailment in *DL-Lite$_{horn}$*   (coNP instead of Π_2^p)

- module extraction algorithm   (extended QBF encoding)

- approximation of Σ-difference