Ontology Learning from Interpretations in Lightweight Description Logics

Szymon Klarman[1] and Katarina Britz[2]

[1] Department of Computer Science, Brunel University London, UK
[2] CSIR Centre for Artificial Intelligence Research, South Africa

August 22, 2015
ILP-15, Kyoto

Ontologies

Ontology: a formal conceptualization of a domain of interest.

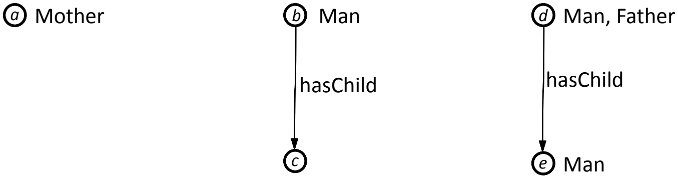
$$\begin{aligned} \mathsf{Mother} &\equiv \mathsf{Woman} \sqcap \exists \mathsf{hasChild}.\top \\ \mathsf{Father} &\equiv \mathsf{Man} \sqcap \exists \mathsf{hasChild}.\top \\ \mathsf{Father\_of\_boy} &\equiv \mathsf{Father} \sqcap \exists \mathsf{hasChild}.\mathsf{Man} \end{aligned}$$

Ontologies

Ontology: a formal conceptualization of a domain of interest.

$$\begin{aligned} \mathsf{Mother} &\equiv \mathsf{Woman} \sqcap \exists \mathsf{hasChild}.\top \\ \mathsf{Father} &\equiv \mathsf{Man} \sqcap \exists \mathsf{hasChild}.\top \\ \mathsf{Father\_of\_boy} &\equiv \mathsf{Father} \sqcap \exists \mathsf{hasChild}.\mathsf{Man} \end{aligned}$$

Instance data + *inference*:

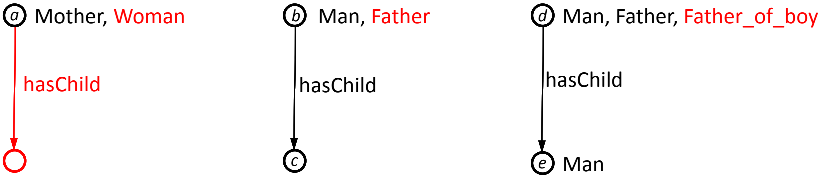


Ontologies

Ontology: a formal conceptualization of a domain of interest.

$$\begin{array}{rcl} \text{Mother} & \equiv & \text{Woman} \sqcap \exists \text{hasChild}.\top \\ \text{Father} & \equiv & \text{Man} \sqcap \exists \text{hasChild}.\top \\ \text{Father\_of\_boy} & \equiv & \text{Father} \sqcap \exists \text{hasChild}.\text{Man} \end{array}$$

Instance data + *inference*:

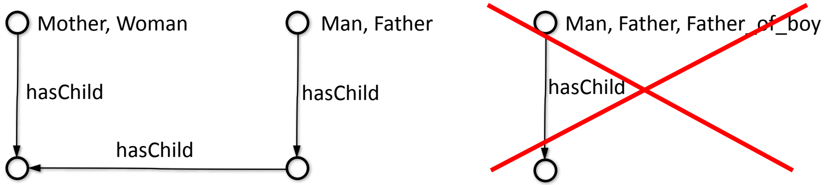


Ontologies

Ontology: a formal conceptualization of a domain of interest.

$$\begin{aligned} \mathsf{Mother} &\equiv \mathsf{Woman} \sqcap \exists \mathsf{hasChild}.\top \\ \mathsf{Father} &\equiv \mathsf{Man} \sqcap \exists \mathsf{hasChild}.\top \\ \mathsf{Father\_of\_boy} &\equiv \mathsf{Father} \sqcap \exists \mathsf{hasChild}.\mathsf{Man} \end{aligned}$$

Models: datasets satisfying all axioms in the ontology:

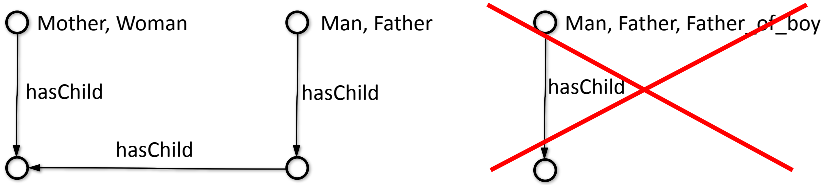


Ontologies

Ontology: a formal conceptualization of a domain of interest.

$$\begin{aligned} \text{Mother} &\equiv \text{Woman} \sqcap \exists \text{hasChild}.\top \\ \text{Father} &\equiv \text{Man} \sqcap \exists \text{hasChild}.\top \\ \text{Father\_of\_boy} &\equiv \text{Father} \sqcap \exists \text{hasChild}.\text{Man} \end{aligned}$$

Models: datasets satisfying all axioms in the ontology:



Can we *learn ontologies from models without supervision*?

Motivation

- *ontologies* and the *Semantic Web technologies* have become a prominent knowledge management paradigm
- the data is abundant and good ontologies are ever more important
- traditional knowledge acquisition bottleneck (building ontologies ex-ante) turns into *knowledge re-engineering bottleneck*:

How to *elicit ontological commitments implicitly present on the data-level*?

Description Logics: syntax

The $\mathcal{EL}$ family of *Description Logics* (underpinning OWL 2 EL) used for modeling large taxonomies (e.g., SNOMED).

Language consists of (atomic) concepts (e.g., Man), roles (e.g., hasChild), and constructors:

$$C, D ::= \overbrace{\underbrace{\top \mid A \mid C \sqcap D}_{\mathcal{L}^{\sqcap}} \mid \exists r.C}^{\mathcal{EL}}$$

An ontology (TBox) is a set of formulas of type: $C \sqsubseteq D$, where:

$\mathcal{EL}$:	C and D in $\mathcal{EL}$
$\mathcal{EL}^{\text{rhs}}$:	C in $\mathcal{L}^{\sqcap}$ and D in $\mathcal{EL}$
$\mathcal{EL}^{\text{lhs}}$:	C in $\mathcal{EL}$ and D in $\mathcal{L}^{\sqcap}$

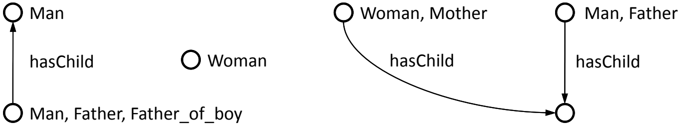
Example:

$\mathcal{EL}^{\text{rhs}}$:	Father_of_boy	$\sqsubseteq$	Man $\sqcap$ $\exists$hasChild.Man
$\mathcal{EL}^{\text{lhs}}$:	Man $\sqcap$ $\exists$hasChild.Man	$\sqsubseteq$	Father_of_boy

Description Logics: semantics

An *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of:

- $\Delta^{\mathcal{I}}$ non-empty domain of individuals,
- interpretation function $\cdot^{\mathcal{I}}$, which encodes a labelled graph, e.g.:



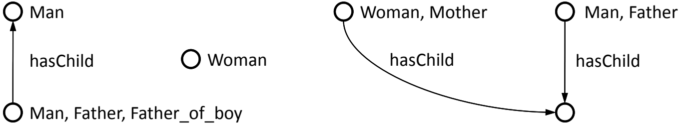
An interpretation $\mathcal{I}$ is a *model* of an ontology *iff* $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for every $C \sqsubseteq D$ in the ontology.

Note: An ontology has infinitely many (possibly) infinite models.

Learning model

Is it possible to *correctly identify an ontology* in an *unsupervised manner* from a *finitely presentable sample of models*?

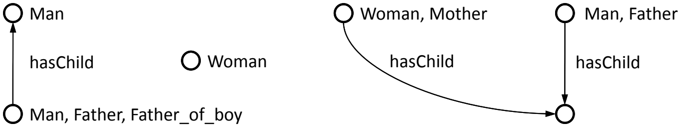
Teacher: provides a finite *learning set*, e.g.:



Learning model

Is it possible to *correctly identify an ontology* in an *unsupervised manner* from a *finitely presentable sample of models*?

Teacher: provides a finite *learning set*, e.g.:



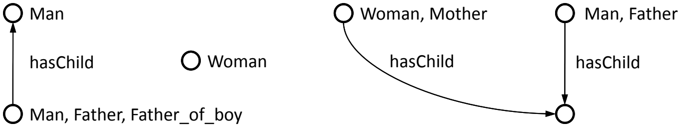
Learner: guesses the *correct ontology*, e.g.:

$$\begin{aligned} \text{Mother} &\equiv \text{Woman} \sqcap \exists \text{hasChild}.\top \\ \text{Father} &\equiv \text{Man} \sqcap \exists \text{hasChild}.\top \\ \text{Father\_of\_boy} &\equiv \text{Father} \sqcap \exists \text{hasChild}.\text{Man} \end{aligned}$$

Learning model

Is it possible to *correctly identify an ontology* in an *unsupervised manner* from a *finitely presentable sample of models*?

Teacher: provides a finite *learning set*, e.g.:



Learner: guesses the *correct ontology*, e.g.:

$$\begin{aligned} \text{Mother} &\equiv \text{Woman} \sqcap \exists \text{hasChild}.\top \\ \text{Father} &\equiv \text{Man} \sqcap \exists \text{hasChild}.\top \\ \text{Father\_of\_boy} &\equiv \text{Father} \sqcap \exists \text{hasChild}.\text{Man} \end{aligned}$$

Admissibility condition: the learning set is reach enough to refute every wrong guess (e.g., there exists a man who is not a woman).

Results

1. There does not exist a finite learning set for $\mathcal{EL}$.
2. There always exist finite learning sets for $\mathcal{EL}^{\text{rhs}}$ and $\mathcal{EL}^{\text{lhs}}$.
   - The learner can succeed easily *using an equivalence oracle*.
   - In case of $\mathcal{EL}^{\text{rhs}}$, the learner can succeed *without the oracle*.

Impossibility result for $\mathcal{EL}$

$$\mathcal{EL}: \quad \exists\mathsf{hasUncle}.\top \quad \sqsubseteq \quad \exists\mathsf{hasParent}.\exists\mathsf{hasSibling}.\mathsf{Man}$$

Theorem: There does not exist a finite learning set for $\mathcal{EL}$.

For every axiom:

$$\underbrace{\exists r.\ldots.\exists r.}_{n}\top \sqsubseteq \underbrace{\exists r.\ldots.\exists r.\exists r.}_{n+1}\top$$

the learning set must contain a finite chain of individuals of length exactly $n+1$, as depicted below:

$$\underbrace{\bullet \xrightarrow{\quad r \quad} \bullet \ldots \bullet \xrightarrow{\quad r \quad} \bullet}_{n+1}$$

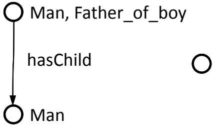
Learning in $\mathcal{EL}^{\text{rhs}}/\mathcal{EL}^{\text{lhs}}$ with an equivalence oracle

$\mathcal{EL}^{\text{rhs}}$: $\mathsf{Father\_of\_boy} \sqsubseteq \mathsf{Man} \sqcap \exists \mathsf{hasChild.Man}$

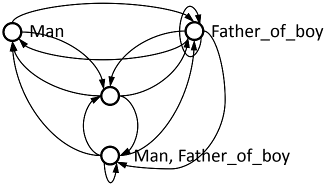
$\mathcal{EL}^{\text{lhs}}$: $\mathsf{Man} \sqcap \exists \mathsf{hasChild.Man} \sqsubseteq \mathsf{Father\_of\_boy}$

Theorem: There always exist finite learning sets for $\mathcal{EL}^{\text{rhs}}$ and $\mathcal{EL}^{\text{lhs}}$ (shown by an application of the type construction method.)

$\mathcal{EL}^{\text{rhs}}$



$\mathcal{EL}^{\text{lhs}}$



Learning in $\mathcal{EL}^{\text{rhs}}/\mathcal{EL}^{\text{lhs}}$ with an equivalence oracle

For every ontology of size n, consistent with the learning set, ask the oracle if it is correct:

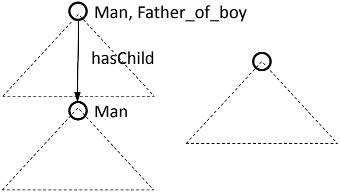
Yes: we're done.

No: increase n by 1.

Unsupervised learning in $\mathcal{EL}^{\text{rhs}}$

$$\mathcal{EL}^{\text{rhs}}: \quad \text{Father\_of\_boy} \sqsubseteq \text{Man} \sqcap \exists \text{hasChild.Man}$$

"Good" candidate axioms can be directly extracted from the learning set in $\mathcal{EL}^{\text{rhs}}$.



The algorithm runs in double exponential time and can generate ontologies of double exponential size w.r.t. data (but there might exist more optimal solutions).

Note: We conjecture a similar result to hold for $\mathcal{EL}^{\text{lhs}}$...

Related work

Ontology learning by query (in Angluin's framework):

Konev, B., Lutz, C., Ozaki, A.,Wolter, F. **Exact learning of lightweight description logic ontologies**. In: *Proc. of KR-14*, 2014.

Konev, B., Ozaki, A.,Wolter, F. **Exact Learning Description Logic Ontologies from Data Retrieval Examples**. In: *Proc. of DL-15*, 2015.

Learning DL concepts via different operators:

Cohen, W.W., Hirsh, H. **Learning the classic description logic: Theoretical and experimental results.**. In: *Proc. of KR-94*, 1994.

Lehmann, J., Hitzler, P. **A refinement operator based learning algorithm for the $\mathcal{ALC}$ description logic.**. In: ILP, 2008.

Applications of concept analysis for enriching ontologies:

Baader, F., Ganter, B., Sertkaya, B., Sattler, U. **Completing description logic knowledge bases using formal concept analysis**. In: *Proc. of IJCAI-07*, 2007.

Conclusions

The presented approach offers a theoretical foundation for the problem of unsupervised ontology learning from data.

Immediate open problem:

- is unsupervised learning possible also in the case of $\mathcal{EL}^{\text{lhs}}$?

Other questions:

- what other feasible conditions could be used to warrant unsupervised learnability in particular languages?
- can we define practical learning algorithms?