Learning Query Inseparable $\mathcal{ELH}$ Ontologies (Extended Abstract)

Ana Ozaki[1], Cosimo Persia[1], and Andrea Mazzullo[2]

[1] University of Bergen, Norway `{ana.ozaki,cosimo.persia}@uib.no`
[2] Free University of Bozen-Bolzano, Italy `mazzullo@inf.unibz.it`

In this extended abstract we give an overview of our work [6], published at AAAI 2020, on learning description logic (DL) ontologies. We focus on learning ontologies formulated as $\mathcal{ELH}$ terminologies [2]. The problem of building ontologies is treated as a learning problem [4, 5, 3] in which a learner (the ontology engineer) attempts to acquire the necessary information to build an ontology by posing queries to a teacher (the domain expert), who knows about the domain of the ontology. This scenario can be modelled within Angluin's exact learning model [1], where a learner interacts with a teacher in order to learn a hypothesis that reflects (by some notion of similarity, approximation or equivalence) an abstract target which, in this case, is a DL ontology.

The learner can formulate questions of two kinds. In the first one, the learner chooses a set of assertions $\mathcal{A}$ and a query q expressible in a query language $\mathcal{Q}$, and asks the teacher if the target ontology $\mathcal{T}$, together with $\mathcal{A}$, entails q.

Example 1. Assume $\mathcal{T} = \{\mathsf{PalmTree} \sqsubseteq \mathsf{TropicalTree}\}$. Then, the learner can ask whether $(\mathcal{T}, \{\mathsf{PalmTree}(\mathsf{a})\}) \models \mathsf{TropicalTree}(\mathsf{a})$, receiving the answer 'yes'.

This is an instance of what is called a *membership query* in Angluin's model (note that the term 'query' here refers to the kind of question posed by the learner, not to a query in a query language). In the second, called *inseparability query*, the learner checks if the hypothesis and the target entail the same queries in a query language $\mathcal{Q}$ on a fixed set of assertions $\mathcal{A}_*$. If so then the hypothesis and the target are *query inseparable* w.r.t. $\mathcal{A}_*$ and $\mathcal{Q}$. In case they are *not* (query) inseparable, the learner receives from the teacher a query in $\mathcal{Q}$ showing that the hypothesis and the target can still be distinguished w.r.t. $\mathcal{A}_*$ and $\mathcal{Q}$. We consider the cases in which $\mathcal{Q}$ is the language of atomic queries AQs, instance queries IQs, conjunctive queries CQs and rooted conjunctive queries CQ_rs (i.e., CQs where each variable is reachable from an individual name in its graph representation).

Example 2. Consider $\mathcal{T} = \{\mathsf{PalmTree} \sqsubseteq \mathsf{TropicalTree}, \mathsf{TropicalTree} \sqsubseteq \mathsf{Tree}\}$, and $\mathcal{H} = \{\mathsf{PalmTree} \sqsubseteq \mathsf{TropicalTree}, \mathsf{PalmTree} \sqsubseteq \mathsf{Tree}\}$. Even though $\mathcal{H}$ is not equivalent to $\mathcal{T}$, it is query inseparable w.r.t. $\mathcal{A}_* = \{\mathsf{PalmTree}(\mathsf{a})\}$ and AQs. However, if $\mathcal{A}_* = \{\mathsf{TropicalTree}(\mathsf{a})\}$ then $\mathcal{T}$ and $\mathcal{H}$ are not inseparable w.r.t $\mathcal{A}_*$ and AQs.

As illustrated in Example 2, query inseparability is a less strict notion than logical equivalence, investigated by Konev et al. [4, 5, 3]. We assume that the

Framework	Equivalent				Inseparable				PAC			
	AQ	IQ	CQ_r	CQ	AQ	IQ	CQ_r	CQ	AQ	IQ	CQ_r	CQ
$\mathcal{EL}(\mathcal{H})_{\mathsf{lhs}}$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\mathcal{EL}(\mathcal{H})_{\mathsf{rhs}}$	−	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓
$\mathcal{EL}(\mathcal{H})$	−	✗	?	✗	✓	✓	✓	✗	✓	✓	✓	✓

Table 1. Polynomial query learnability of learning frameworks.

learner knows the vocabulary of the ontology about the domain of interest (but not how terms relate to each other and how they can be combined to form the ontology), and that the teacher answers accordingly. We study, in the model just described, the complexity of learning query inseparable $\mathcal{ELH}$ ontologies.

The complexity of learning in this case depends on the interactions between the learner and the teacher in order to learn a query inseparable ontology, where posing a (membership or inseparability) query counts as one step of computation. We consider two notions of complexity. In the case of query inseparability, they depend also on the size of the fixed ABox, which is considered to be part of input. One is based on the number and size of queries posed by the learner. This is the *query* complexity. The other is based on the overall time spent by the learner—*time* complexity (for learning). Polynomial time learnability implies polynomial query learnability but not the converse.

However, answering inseparability queries is more challenging for an expert than answering membership queries. For this reason, we also consider an approach for learning ontologies, based on Valiant's probably approximately correct (PAC) model [7] that avoids inseparability queries. Table 1 summarises, in the first two columns, polynomial query learnability results for finding a logically equivalent hypothesis, and our query inseparability results. We also consider the cases where ontologies contain CIs with complex concepts only on the left hand side ($\mathcal{EL}(\mathcal{H})_{\mathsf{lhs}}$), or only on the right hand side ($\mathcal{EL}(\mathcal{H})_{\mathsf{rhs}}$). The symbol ✓ means 'polynomial query learnable', while ✗ means the negation of ✓. The last column of the table shows that query inseparable ontologies are polynomial query learnable in the PAC model. Our main contributions are shaded in grey. In addition, we show that there is a case in which one can achieve PAC learnability in polynomial time but not polynomial query learnability with membership and inseparability queries (and thus also not polynomial time learnability).

Query inseparability may not be preserved after that the predefined set of assertions is updated (Example 3).

Example 3. If the fixed set of assertions in Example 2 is updated to $\mathcal{A}_* = \{\mathsf{TropicalTree}(\mathsf{a})\}$, then $\mathcal{H}$ and $\mathcal{T}$ of Example 2 are not anymore query inseparable w.r.t. $\mathcal{A}_*$ and AQ, because $(\mathcal{T}, \mathcal{A}_*) \models \mathsf{Tree}(\mathsf{a})$, while $\mathcal{H}$ does not.

For this reason, we also study conditions under which query inseparability is guaranteed to be preserved. We first observe that, if every individual in the updated data instance $\mathcal{A}$ is *bisimilar* to an individual in the fixed data instance $\mathcal{A}_*$, then IQ-inseparability is preserved. This result does not hold if we just

require the existence of a homomorphism from the old ABox to the updated one and vice-versa. We then show that there is an algorithm that can compute a hypothesis that would preserve IQ inseparability under an extended class of possible updates while still preserving IQ inseparability. The update in Example 3 is covered by this extended class of updates even though there is no bisimulation between the old and the new set of assertions. In case query inseparability is not preserved, the learning algorithm can pose further queries starting from the previously built hypothesis.

References

1. Dana Angluin. Queries and concept learning. *Machine Learning*, 2(4):319–342, 1988.
2. Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, second edition, 2007.
3. Mario Ricardo Cruz Duarte, Boris Konev, and Ana Ozaki. ExactLearner: A tool for exact learning of $\mathcal{EL}$ ontologies. In *KR*, pages 409–414, 2018.
4. Boris Konev, Carsten Lutz, Ana Ozaki, and Frank Wolter. Exact Learning of Lightweight Description Logic Ontologies. *Journal of Machine Learning Research*, 18(201):1–63, 2018.
5. Boris Konev, Ana Ozaki, and Frank Wolter. A model for learning description logic ontologies based on exact learning. In *AAAI*, pages 1008–1015, 2016.
6. Ana Ozaki, Cosimo Persia, and Andrea Mazzullo. Learning query inseparable $\mathcal{ELH}$ ontologies. In *AAAI*, pages 2959–2966. AAAI Press, 2020.
7. L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.