Learning Ontologies with Epistemic Reasoning: The $\mathcal{EL}$ Case (Extended Abstract)

Ana Ozaki and Nicolas Troquard

KRDB Research Centre, Free University of Bozen-Bolzano, Italy
{ana.ozaki,nicolas.troquard}@unibz.it

This extended abstract gives an overview of the work presented in [8]. We are interested in the formal and computational aspects of the problem of an agent learning from the knowledge of another agent, playing the role of a domain expert. In general, the raw knowledge of an agent is not directly accessible. In particular, the knowledge of an expert, the *target* domain of interest, cannot be duplicated as is and transferred to a learner. Nonetheless, agents can still learn from one another through questions and answers in a communication protocol. These acts of communication are expected to be simple.

A classical communication protocol from computational learning theory is based on questions of two types: *membership* and *equivalence* queries [1]. In a learning from entailments setting [5], these questions can be described as follows. Membership queries correspond to asking whether a certain statement formulated as a logical sentence follows from the target. Equivalence queries correspond to asking whether a certain logical theory, called *hypothesis*, precisely describes the target. If there are wrong or missing statements in the hypothesis, a statement illustrating the imprecision should be returned to the agent playing the role of the *learner*. Arguably, equivalence queries are *not* in fact simple, and this is one of the main difficulties in implementing this protocol in practice [7, page 297]. Whenever a learner poses an equivalence query, the expert playing the role of an *oracle* needs to evaluate the whole hypothesis and decide whether or not it is equivalent to the target. If not, then the oracle returns a statement in the logical difference between the hypothesis and the target.

One way out of this difficulty is hinted to us by a simple observation: during interactive communication among agents, not only domain knowledge is exchanged and acquired but also second-order knowledge, which is the knowledge of what is known by the other agents. We propose a new and more realistic learning model which takes into account what is known by the agents, either because a statement was explicitly communicated or because it is a logical consequence of previous statements given during their interaction. Our protocol is based on queries of two types. The first is an epistemic version of membership queries where the oracle 'remembers' those membership queries whose reply was 'yes'. We call the second type example queries. When asked an example query, the oracle answers a statement which follows from its knowledge but does not follow from its knowledge about what the learner knows. The oracle also 'remembers' that the statements given are now known by the learner.

Formally, each $i \in \mathbf{A}$ aims at learning a *target* formula $l_j \in L_j$ of each other agent $j \neq i \in \mathbf{A}$ by posing them queries. A membership query to i asks an oracle

$$\text{EPISTEMIC}[\mathsf{MEM},\mathsf{EX}] \overset{[8,\ \text{Th. } 3]}{=} \text{EXACT}[\mathsf{MEM},\mathsf{EQ}] \overset{[7] \text{ and } [2]}{\subset} \text{PAC}[\mathsf{MEM}]$$

$$\text{EPISTEMIC}[\mathsf{EX}] \overset{[8,\ \text{Th. } 1 \text{ and } 2]}{\subset} \text{EXACT}[\mathsf{EQ}] \overset{[1] \text{ and } [2]}{\subset} \text{PAC}$$

Fig. 1. Polynomial learnability. Each class denotes the set of frameworks that are polynomial query learnable in the corresponding learning model. MEM, EQ and EX stand for membership, equivalence, and example queries respectively.

MEM whether an example $x \in X_i$ is entailed by l_i. An equivalence query to i asks an oracle EQ whether a formula $h \in L_i$ is equivalent to l_i; it returns yes if it is the case, or an counterexample in X_i otherwise. Here, we introduce new kinds of queries. A **K**-membership query to i asks an oracle $\mathsf{MEM}^{\mathsf{K}}$ whether an example $x \in X_i$ is entailed by l_i. When it is the case, the oracle $\mathsf{MEM}^{\mathsf{K}}$ also remembers that the learner j knows that x, adding $\mathbf{K}_j x$ to its own knowledge. An example query to i requests the oracle EX to provide an example x that is entailed by l_i but does such that $\mathbf{K}_j x$ does not follow from the oracle's knowledge. An *exact learning algorithm* for a learning framework is a deterministic algorithm that takes no input, is allowed to make queries to MEM and EQ, and eventually halts and outputs some $h \in L_i$, equivalent to l_i. An *epistemic learning algorithm* is similar to an exact learning algorithm, except that it is only allowed to make queries to $\mathsf{MEM}^{\mathsf{K}}$ and EX. Polynomial time learnability means that there is an algorithm such that the time used is bounded by a polynomial in the size of the target and the largest counterexample seen so far. An analogous notion depending on the size and number of queries is used for polynomial query learnability [8].

We show that if a multi-agent learning framework is polynomial query (resp. time) epistemically learnable with only example queries then it is polynomial query (resp. time) exactly learnable with only equivalence queries; while the other direction does not hold. Example queries to EX are indeed less powerful than equivalence queries to EQ. Indeed, we show that if a multi-agent learning framework is polynomial query (resp. time) epistemically learnable with only example queries then it is polynomial query (resp. time) exactly learnable with only equivalence queries. This is summarized in Figure 1, together with known results about the relationship between exact learning and PAC learning. We then instantiate our learning framework to $\mathcal{EL}$. We prove that satisfiability of conjunctive formulas of $\mathcal{EL}$ extended with epistemic modal operators (called conjunctive $\mathcal{ELK}$) is PTime-complete (see, [8, Th. 5]). Conjunctive $\mathcal{ELK}$ captures the expressivity used in the epistemic learning model. Together with the results in Figure 1, the PTime complexity of the satisfiability problem of conjunctive $\mathcal{ELK}$ demonstrates that epistemic learning is a reasonable substitute to the exact learning model in the $\mathcal{EL}$ case. Finally, we transfer known results [3,4,6] for exact learnability of $\mathcal{EL}$ ontologies and its fragments to our learning model.

References

1. Angluin, D.: Queries and concept learning. Machine Learning **2**(4), 319–342 (1988)
2. Blum, A.L.: Separating distribution-free and mistake-bound learning models over the boolean domain. SIAM J. Comput. **23**(5) (1994)
3. Duarte, M.R.C., Konev, B., Ozaki, A.: Exact learning of EL ontologies. In: Proceedings of the 31st International Workshop on Description Logics co-located with 16th International Conference on Principles of Knowledge Representation and Reasoning (KR) (2018)
4. Duarte, M.R.C., Konev, B., Ozaki, A.: Exactlearner: A tool for exact learning of EL ontologies. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR. pp. 409–414 (2018)
5. Frazier, M., Pitt, L.: Learning from entailment: An application to propositional Horn sentences. In: International Conference on Machine Learning, ICML. pp. 120–127 (1993)
6. Konev, B., Lutz, C., Ozaki, A., Wolter, F.: Exact learning of lightweight description logic ontologies. Journal of Machine Learning Research **18**(201), 1–63 (2018)
7. Mohri, M., Rostamizadeh, A., Talwalkar, A.: Foundations of Machine Learning. The MIT Press (2012)
8. Ozaki, A., Troquard, N.: Learning Ontologies with Epistemic Reasoning: The EL Case. In: Proceedings of the 16th European Conference on Logics in Artificial Intelligence (JELIA 2019) (2019), to appear