

PAC Learning of Concept Inclusions for Ontology-Mediated Query Answering (Extended Abstract)

Sergei Obiedkov¹, Barış Sertkaya²

¹Faculty of Computer Science / cfaed / ScaDS.AI, TU Dresden, Dresden, Germany

²Faculty of Computer Science and Engineering, Frankfurt University of Applied Sciences, Germany

Keywords

Ontologies, Description Logics, Active Learning, Knowledge Acquisition, PAC Learning

1. Introduction

This extended abstract summarizes results from [1], where we propose a practical method for learning axioms in a Description Logic (DL) ontology using techniques from probably approximately correct (PAC) learning. The goal is to support ontology-mediated query answering (OMQA) [2, 3] by approximating an unknown TBox \mathcal{T} through interaction with a *domain expert oracle* that can decide whether a concept inclusion (CI) $C \sqsubseteq D$ is entailed by \mathcal{T} . Such an oracle may be instantiated in different ways—for example, as a human domain expert; a large language model (LLM); a dataset representative of the domain; or a large, complex ontology from which a smaller, focused one is to be distilled.

Our method learns subsumption relationships among a finite set \mathcal{C} of concept descriptions, called the *base set*. This base set constrains the search space of candidate axioms and can be chosen to suit the application—e.g., all concept names, combinations of concept names with existential restrictions up to a fixed role depth, or a tailored selection relevant to the user. We do not fix a particular DL or a set of constructors; our results apply to arbitrary DLs that support conjunction.

The algorithm also employs a *sampling oracle* that generates CIs over \mathcal{C} according to a fixed but arbitrary distribution \mathcal{D} . Given $\epsilon, \delta \in (0, 1)$, it runs in time polynomial in the relevant parameters and returns a TBox \mathcal{T}' such that, with probability at least $1 - \delta$ (over the algorithm's random choices), the probability (under \mathcal{D}) that a CI over \mathcal{C} is entailed by exactly one of \mathcal{T} and \mathcal{T}' is at most ϵ .

We also show how to direct the learning process toward subsumptions that are particularly relevant to a given ABox \mathcal{A} , by adapting the distribution \mathcal{D} . This enables the learned axioms to improve recall in query answering over incomplete datasets. Experimental evaluation on benchmark ontologies confirms the effectiveness of our approach.

For related work on ontology learning in DLs, see [4, 5, 6, 7, 8, 9, 10, 11, 12].

2. PAC Learning of Concept Inclusions

Let \mathcal{T} be a TBox, \mathcal{C} be a *base set* of concept descriptions over its signature, and \mathcal{D} be a probability distribution on CIs of the form $\bigwedge \mathcal{X} \sqsubseteq D$, where $\mathcal{X} \subseteq \mathcal{C}$ and $D \in \mathcal{C}$. For $0 < \epsilon < 1$, we say that a set \mathcal{T}' of such CIs is an ϵ - \mathcal{C} -approximation of \mathcal{T} if

$$\Pr_{\mathcal{D}}(q \mid (\mathcal{T} \models q) \iff (\mathcal{T}' \not\models q)) \leq \epsilon,$$

where q is such a CI. \mathcal{T}' is called a *lower ϵ - \mathcal{C} -approximation* if $\mathcal{T} \models \mathcal{T}'$.

We base our solution on an algorithm for exactly learning propositional Horn formulas [13], which requires two types of queries: membership and equivalence queries. We simulate membership queries

using subsumption queries. We modify the equivalence oracle so that, instead of returning a model of exactly one of the two non-equivalent Horn formulas, it returns the GCI corresponding to an Horn clause that is entailed by exactly one of the two formulas. A probably approximately correct algorithm is obtained by replacing each call to this equivalence oracle with an appropriate number of calls to a suitable sampling oracle. Please see [1] for more details.

3. Varying the Query Distribution

Our definition of approximation involves a distribution \mathcal{D} of subsumption queries. This distribution is meant to reflect the interests of the user of the ontology we are trying to learn. In a basic scenario, we may assume that users explicitly pose subsumption queries to the ontology and that \mathcal{D} is the distribution of these queries.

A more practically relevant scenario is given by ontology-mediated query answering [2, 3]: Given a query (a concept description) q and a knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, find all instances of q in \mathcal{K} . The TBox \mathcal{T} may be only partially known or not known at all, in which case we may use our PAC algorithm to learn its approximation through interaction with an expert or from a representative dataset. In this setting, an approximation may be considered good if it ensures high precision and recall of query answering.

Definition 1. Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a knowledge base, \mathcal{T}' be a TBox, and q be a query. Using certain-answer semantics [2], we define $\text{cert}(q, \mathcal{K})$ as the set of individual names a from \mathcal{A} satisfying $\mathcal{K} \models q(a)$. The precision and recall of \mathcal{T}' for q on \mathcal{K} are, respectively,

$$P^q(\mathcal{T}', \mathcal{K}) = \frac{|\text{cert}(q, (\mathcal{T}', \mathcal{A})) \cap \text{cert}(q, \mathcal{K})|}{|\text{cert}(q, (\mathcal{T}', \mathcal{A}))|}, R^q(\mathcal{T}', \mathcal{K}) = \frac{|\text{cert}(q, (\mathcal{T}', \mathcal{A})) \cap \text{cert}(q, \mathcal{K})|}{|\text{cert}(q, \mathcal{K})|}.$$

If the denominator is 0, then the value of the corresponding measure is defined to be 1.

There are two standard ways to aggregate precision and recall for several queries: macroaveraging and microaveraging [14].

Definition 2. Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a knowledge base, \mathcal{T}' be a TBox, and Q be a finite set of queries. The macro precision and recall of \mathcal{T}' for Q on \mathcal{K} are the average values of the precision and recall over all queries from Q :

$$P_{\text{macro}}^Q(\mathcal{T}', \mathcal{K}) = \frac{\sum_{q \in Q} P^q(\mathcal{T}', \mathcal{K})}{|Q|} \quad \text{and} \quad R_{\text{macro}}^Q(\mathcal{T}', \mathcal{K}) = \frac{\sum_{q \in Q} R^q(\mathcal{T}', \mathcal{K})}{|Q|}.$$

The micro precision $P_{\text{micro}}^Q(\mathcal{T}', \mathcal{K})$ and micro recall $R_{\text{micro}}^Q(\mathcal{T}', \mathcal{K})$ are defined, respectively, as

$$\frac{\sum_{q \in Q} |\text{cert}(q, (\mathcal{T}, \mathcal{A})) \cap \text{cert}(q, (\mathcal{T}', \mathcal{A}))|}{\sum_{q \in Q} |\text{cert}(q, (\mathcal{T}', \mathcal{A}))|} \quad \text{and} \quad \frac{\sum_{q \in Q} |\text{cert}(q, (\mathcal{T}, \mathcal{A})) \cap \text{cert}(q, (\mathcal{T}', \mathcal{A}))|}{\sum_{q \in Q} |\text{cert}(q, (\mathcal{T}, \mathcal{A}))|}.$$

The goal in our OMQA scenario is to learn an approximation \mathcal{T}' of \mathcal{T} with high values of the macro/micro precision and recall for some set Q of queries. If \mathcal{T}' is a lower approximation of \mathcal{T} , then the precision for every query is 1, and so are the macro and micro precision. In this case, we aim to maximize the recall. Next we describe a heuristic approach to choosing the distribution of subsumption queries in the learning algorithm so as to increase the micro recall on a given ABox \mathcal{A} .

Consider a subsumption query $\sqcap \mathcal{X} \sqsubseteq D$. If we care about micro recall, it seems particularly important to ask this query whenever $\sqcap \mathcal{X}$ has a lot of instances in $\mathcal{K}_0 = (\emptyset, \mathcal{A})$, since a positive answer to the query would then allow us to correctly assert $D(x)$ for many individuals x . Therefore, a reasonable approach seems to be to generate the left-hand sides $\sqcap \mathcal{X}$ of subsumption queries proportionally to $|\text{cert}(\sqcap \mathcal{X}, \mathcal{K}_0)|$. Regarding the right-hand sides, if $D(x)$ rarely occurs in \mathcal{A} , this may be due to two reasons: D is a rare concept, or D is a generalization of other concepts and $D(x)$ can be inferred

Table 1 ϵ and lower ϵ -approximations for the ore_ont_5596 ontology

	ϵ	ϵ -approximation			lower ϵ -approximation			
		\mathcal{A} -induced			uniform		\mathcal{A} -induced	
		Macro P/R	Micro P/R	$ \mathcal{T}' $	Macro/Micro R	$ \mathcal{T}' $	Macro/Micro R	$ \mathcal{T}' $
\mathcal{C}_1	–	1.0 / 0.57	1.0 / 0.10	0				
	0.1	1.0 / 0.75	1.0 / 0.70	7.4	0.71 / 0.62	24.4	0.75/0.70	6.6
	0.01	0.99 / 0.98	0.99 / 0.99	24.8	0.78 / 0.69	30.6	0.97/0.98	21.2
	0.005	1.0 / 0.98	1.0 / 0.99	26	0.76 / 0.71	33	0.98/0.99	25.8
\mathcal{C}_2	–	1.0 / 0.73	1.0 / 0.36	0				
	0.1	0.98 / 0.84	0.97 / 0.83	7.4	0.85 / 0.80	28.4	0.82/0.80	6.2
	0.01	0.99 / 0.98	0.99 / 0.99	30.8	0.85 / 0.81	38	0.97 / 0.98	29.4
	0.005	0.99 / 0.99	0.99 / 0.99	37.8	0.87 / 0.84	41	0.99 / 0.99	34.8

from the target TBox \mathcal{T} together with what is explicitly asserted in \mathcal{A} about x . We cannot tell which of the two it is; so we may want to assume the second case to be on the safe side. Then, we may want to generate the right-hand sides D of subsumption queries with probabilities proportional to $|\text{Ind}(\mathcal{A}) \setminus \text{cert}(D, \mathcal{K}_0)|$, i.e., to the number of individuals that are not (yet) known to be instances of D .

A problem with this approach is that $B \sqsubseteq C$ cannot be learned if B has no instances in \mathcal{K}_0 . To address this, we need to change the distribution on the fly, so as to take into account what has already been learned. Thus, having learned $A \sqsubseteq B$, we update \mathcal{K}_0 by replacing $\mathcal{T}_0 = \emptyset$ with $\mathcal{T}_1 = \{A \sqsubseteq B\}$ and recalculate the probabilities involved in sampling premises with respect to $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$. Now, $|\text{cert}(B, \mathcal{K}_1)| > 0$, which makes it possible to learn $B \sqsubseteq C$.

This was the method used in the experiments we presented in [1]. However, it prioritizes concepts $\sqcap \mathcal{X}$ with a large number of instances in \mathcal{A} even when these instances are the same for many different \mathcal{X} . This may sometimes negatively affect precision or recall for certain concepts in \mathcal{C} . Instead, when sampling left-hand sides of CIs, we should try to maximize the coverage of individuals in \mathcal{A} . Therefore, in the experiments presented here, we adopt the following two-stage approach: first sample an individual a from \mathcal{A} uniformly at random and then sample a subset of $\{C \in \mathcal{C} \mid a \in \text{cert}(C, \mathcal{K}_i)\}$ also uniformly at random.

4. Experimental Evaluation

We implemented our approach in a prototype tool, PACLO¹, and evaluated it in the OMQA context. The expert was simulated using the target TBox \mathcal{T} ; i.e., the response to a subsumption query q is positive if and only if $\mathcal{T} \models q$. Subsumption queries were answered with the ELK reasoner [15]. We set $\delta = 0.001$ to ensure a high probability of obtaining the desired approximation and averaged results over five runs. Each setting is defined by a signature, an approximation type (ϵ - or lower ϵ -approximation), and a query distribution \mathcal{D} (uniform or \mathcal{A} -induced, as described in the previous section).

We tested on six KBs: four generated with OWL2Bench [16] and two from the ORE 2015 repository [17]. Due to space constraints, we report results here only for the KB *ore_ont_5596*; results for the other KBs can be found in [1]. The *ore_ont_5596* KB contains 58 class names, 33 role names, 322 GCIs, 112,320 individuals, 32,990 class assertions and 190,149 role assertions. Two base sets were considered: \mathcal{C}_1 with all class names and \mathcal{C}_2 that adds $\exists r. \top$ for each role r , yielding 93 concepts in total. We measure the macro and micro recall on the query set $Q = \mathcal{C}_i$. The results are shown in Table 1.

For each of \mathcal{C}_1 and \mathcal{C}_2 , the first line represents the precision and recall of the empty \mathcal{T}' , i.e., the quality of query answering based only on the ABox. This serves our baseline. We omit ϵ -approximations with the uniform distribution, since it provides hardly any improvement over the baseline. The best macro and micro recall values for each ϵ are shown in bold. The column $|\mathcal{T}'|$ contains the number of axioms learned.

¹<https://github.com/sertkaya/paclo>

Overall, \mathcal{A} -induced distributions yield substantially higher recall, particularly for micro recall. Lower approximations typically have slightly reduced recall, but may be preferable when perfect precision is required. Lower approximations for the uniform distribution do show some improvement over the baseline, but usually smaller than those for the \mathcal{A} -induced distribution and with larger sets of GCIs. Note that the perfect recall may not be achievable with a fixed base set \mathcal{C} , since \mathcal{T} may contain axioms mixing concepts from and outside \mathcal{C} .

Acknowledgments

This work is partly supported by DFG in project 389792660 (TRR 248, Center for Perspicuous Systems), by BMBF in ScaDS.AI, and by BMBF and DAAD in project 57616814 (SECAI).

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to: Improve writing style. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] S. Obiedkov, B. Sertkaya, PAC learning of concept inclusions for ontology-mediated query answering, *International Journal of Approximate Reasoning* 186 (2025) 109523. URL: <https://www.sciencedirect.com/science/article/pii/S0888613X25001641>. doi:<https://doi.org/10.1016/j.ijar.2025.109523>.
- [2] M. Bienvenu, M. Ortiz, Ontology-mediated query answering with data-tractable description logics, in: W. Faber, A. Paschke (Eds.), *Reasoning Web. Web Logic Rules - 11th International Summer School 2015, Berlin, Germany, July 31 - August 4, 2015, Tutorial Lectures*, volume 9203 of *Lecture Notes in Computer Science*, Springer, 2015, pp. 218–307. URL: https://doi.org/10.1007/978-3-319-21768-0_9. doi:10.1007/978-3-319-21768-0_9.
- [3] M. Bienvenu, Ontology-mediated query answering: Harnessing knowledge to get more from data, in: S. Kambhampati (Ed.), *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016, IJCAI/AAAI Press*, 2016, pp. 4058–4061. URL: <http://www.ijcai.org/Abstract/16/600>.
- [4] F. Baader, B. Ganter, B. Sertkaya, U. Sattler, Completing description logic knowledge bases using formal concept analysis, in: M. M. Veloso (Ed.), *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI'07)*, AAAI Press, 2007, pp. 230–235.
- [5] D. Borchmann, F. Distel, F. Kriegel, Axiomatisation of general concept inclusions from finite interpretations, *Journal of Applied Non-Classical Logics* 26 (2016) 1–46. URL: <https://doi.org/10.1080/11663081.2016.1168230>. doi:10.1080/11663081.2016.1168230. arXiv:<https://doi.org/10.1080/11663081.2016.1168230>.
- [6] B. Konev, C. Lutz, A. Ozaki, F. Wolter, Exact learning of lightweight description logic ontologies, *Journal of Machine Learning Research* 18 (2018) 1–63. URL: <http://jmlr.org/papers/v18/16-256.html>.
- [7] A. Ozaki, On the complexity of learning description logic ontologies, in: M. Manna, A. Pieris (Eds.), *Reasoning Web. Declarative Artificial Intelligence: 16th International Summer School 2020, Oslo, Norway, June 24–26, 2020, Tutorial Lectures*, Springer International Publishing, Cham, 2020, pp. 36–52. URL: https://doi.org/10.1007/978-3-030-60067-9_2. doi:10.1007/978-3-030-60067-9_2.
- [8] A. Ozaki, C. Persia, A. Mazzullo, Learning query inseparable \mathcal{ELH} ontologies, *Proceedings of the AAAI Conference on Artificial Intelligence* 34 (2020) 2959–2966. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/5688>. doi:10.1609/aaai.v34i03.5688.
- [9] M. Funk, J. C. Jung, C. Lutz, Actively learning concepts and conjunctive queries under elr-ontologies, in: Z. Zhou (Ed.), *Proceedings of the Thirtieth International Joint Conference on*

Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021, ijcai.org, 2021, pp. 1887–1893. URL: <https://doi.org/10.24963/ijcai.2021/260>. doi:10.24963/IJCAI.2021/260.

- [10] B. ten Cate, M. Funk, J. C. Jung, C. Lutz, SAT-based PAC learning of description logic concepts, in: E. Elkind (Ed.), Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23, International Joint Conferences on Artificial Intelligence Organization, 2023, pp. 3347–3355. URL: <https://doi.org/10.24963/ijcai.2023/373>. doi:10.24963/ijcai.2023/373, main Track.
- [11] M. Funk, Learning queries under description logic ontologies, Ph.D. thesis, Universität Leipzig, 2025.
- [12] F. Kriegel, Efficient axiomatization of owl 2 el ontologies from data by means of formal concept analysis, Proceedings of the AAAI Conference on Artificial Intelligence 38 (2024) 10597–10606. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/28930>. doi:10.1609/aaai.v38i9.28930.
- [13] D. Angluin, M. Frazier, L. Pitt, Learning conjunctions of horn clauses, Machine Learning 9 (1992) 147–164. URL: <https://link.springer.com/article/10.1007/bf00992675>. doi:<https://doi.org/10.1007/BF00992675>.
- [14] C. D. Manning, P. Raghavan, H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.
- [15] Y. Kazakov, M. Krötzsch, F. Simancik, The incredible ELK - from polynomial procedures to efficient reasoning with \mathcal{EL} ontologies, J. Autom. Reason. 53 (2014) 1–61. URL: <https://doi.org/10.1007/s10817-013-9296-3>. doi:10.1007/s10817-013-9296-3.
- [16] G. Singh, S. Bhatia, R. Mutharaju, Owl2bench: A benchmark for OWL 2 reasoners, in: The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part II, volume 12507 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 81–96.
- [17] M. Dumontier, B. Glimm, R. S. Gonçalves, M. Horridge, E. Jiménez-Ruiz, N. Matentzoglou, B. Parsia, G. B. Stamou, G. Stoilos (Eds.), Informal Proceedings of the 4th International Workshop on OWL Reasoner Evaluation (ORE-2015) co-located with the 28th International Workshop on Description Logics (DL 2015), Athens, Greece, June 6, 2015, volume 1387 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2015.