Supplemental Material

Anonymous

1 Proof for soundness

Proposition 1 *(Soundness of the semantics) Let $\sharp\mathcal{O}$ be an $\mathcal{ALC}$ ontology, and φ be a fuzzy assertion. $\mathcal{O} \approx\!\!\!\!\models \varphi$ iff. $\sharp\mathcal{O} \models \sharp\varphi$ (i.e. fuzzy entailment is consistent with entailment in $\mathcal{ALC}$).*

Proof. 1.$\Rightarrow$ Consider any crisp interpretation $\mathcal{I}$ that is a model of $\sharp\mathcal{O}$. $\mathcal{I}$ can also be considered as a fuzzy interpretation that $C^{\mathcal{I}}(a) \in \{0, 0.5, 1\}$ and $r^{\mathcal{I}}(a, b) \in \{0, 0.5, 1\}$ hold. By induction on the structure of a concept C, $\mathcal{I}$ satisfies $a : C$ iff $C^{\mathcal{I}}(a) = 1$. $\mathcal{I}$ satisfies $a : C$ is unknown iff $C^{\mathcal{I}}(a) = 0.5$. And similarly for roles. Therefore, $\mathcal{I}$ is also a model of $\mathcal{O}$. And for every model of $\sharp\mathcal{O}$, $\mathcal{I}$ satisfies $\sharp\varphi$, so $\sharp\mathcal{O} \models \sharp\varphi$ holds.

2. $\Leftarrow$ If $\sharp\mathcal{O} \models \sharp\varphi$, consider the crisp interpretation $\mathcal{I}$ discussed above, it is similar to proof that each model $\mathcal{I}$ of $\sharp\mathcal{O}$, is also the model of $\mathcal{O}$, and satisfies φ, so we have $\mathcal{O} \approx\!\!\!\!\models \varphi$. To sum up, this proposition is proven to be true.

Theorem 1. *For any $\mathcal{ALC}$ ontology $\mathcal{O}$, one can construct in polynomial time a normalized $\mathcal{ALC}$-ontology $\mathcal{O}$' of polynomial size in $|\mathcal{O}|$ using the normalization described above such that (i) for every model $\mathcal{I}$ of $\mathcal{O}$, there exists a model $\mathcal{J}$ of $\mathcal{O}'$ such that $\mathcal{I}$ is semantically equivalent to $\mathcal{J}$ in $sig(O)$, denoted as $\mathcal{I} \sim_{sig(O)} \mathcal{J}$, and (ii) for every model $\mathcal{J}$ of $\mathcal{O}'$ there exists a model $\mathcal{I}$ of $\mathcal{O}$ such that $\mathcal{I} \sim_{sig(O)} \mathcal{J}$.*

Proposition 2 *(**Soundness of learning to ground in DF-**$\mathcal{ALC}$) When the hierarchical loss converges to 0, the learned interpretation $\mathcal{I}''$ is the model of the given $\mathcal{ALC}$ ontology $\mathcal{O}$. For any model $\mathcal{J}$ of $\mathcal{O}$, $\mathcal{I}'' \sim_{sig(\mathcal{O})} \mathcal{J}$.*

Proof. when loss converges to 0, the learned $\mathcal{I}''$ satisfies any $C \sqsubseteq D$ in the normalized ontology $\mathcal{O}'$, so $\mathcal{I}''$ is the model of $\mathcal{O}'$. And according to Theorem 1, any model of $\mathcal{O}'$ is semantically equivalent to the model of $\mathcal{O}$. So this proposition is proved to be true.

2 Rule-based Learning: Example Analyse

Example 1. Given a perceptual grounding $\mathcal{I}$ in the domain $\{s_1, s_2\}$, $\mathtt{A}^{\mathcal{I}}(s_1) = 0, \mathtt{A}^{\mathcal{I}}(s_2) = 0, \mathtt{B}^{\mathcal{I}}(s_1) = 0.9, \mathtt{B}^{\mathcal{I}}(s_2) = 0, \mathtt{r}^{\mathcal{I}}(s_1, s_2) = 0.9, \mathtt{r}^{\mathcal{I}}(s_1, s_1) = \mathtt{r}^{\mathcal{I}}(s_2, s_1) = \mathtt{r}^{\mathcal{I}}(s_2, s_2) = 0$, notated as vectors $\mathtt{A}^{\mathcal{I}} = [0, 0]$, $\mathtt{B}^{\mathcal{I}} = [0.9, 0]$.

According to the semantics of fuzzy $\mathcal{ALC}$, in $\mathcal{O}_1$, $(\exists \mathtt{r}.\mathtt{A})^{\mathcal{I}} = [0, 0]$, which satisfies $(\exists \mathtt{r}.\mathtt{A})^{\mathcal{I}} \leq \mathtt{B}^{\mathcal{I}}$, so hierarchical loss is 0, and no revision is executed. But this is not what we want. As we know that s_1 is likely to be $\mathtt{B}$, and $\mathtt{r}(s1, s2)$ is likely to be true, so s_2 is likely to be a membership of $\mathtt{A}$. In $\mathcal{O}_2$, $(\forall \mathtt{r}.\mathtt{A})^{\mathcal{I}} = [0.1, 1]$, which does not satisfy $(\forall \mathtt{r}.\mathtt{A})^{\mathcal{I}} \leq \mathtt{B}^{\mathcal{I}}$, and hierarchical loss is 1.1. Through gradient decent, until loss becomes 0, $\mathtt{A}^{\mathcal{I}} = [0.24, 0]$,$\mathtt{B}^{\mathcal{I}} = [0.4, 1]$, $\mathtt{r}^{\mathcal{I}}(s_2, s_1)$ will be 0.7 and $\mathtt{r}^{\mathcal{I}}(s_1, s_2)$ will be 1. In $\mathcal{O}_3$, with hierarchical loss, $\mathcal{A}^{\mathcal{I}}$ will be $[0.38, 0]$, $\mathcal{B}^{\mathcal{I}}$ will be $[0.35, 0]$and $\mathtt{r}^{\mathcal{I}}(s_1, s_1)$ will be $[0, 36]$. In $\mathcal{O}_4$, with hierarchical loss, $\mathcal{B}^{\mathcal{I}} = [0, 0]$ and $\mathtt{r}^{\mathcal{I}}(s_1, s_2) = 0$.

Example 2. Given a perceptual grounding $\mathcal{I}$ in the domain $\{s_1, s_2\}$, $\mathtt{A}^{\mathcal{I}} = [0, 0.9]$, $\mathtt{B}^{\mathcal{I}} = [0, 0]$, $\mathtt{r}^{\mathcal{I}}$ is the same as in Example. 1.

According to the semantics of fuzzy $\mathcal{ALC}$, in $\mathcal{O}_1$, $(\exists \mathtt{r}.\mathtt{A})^{\mathcal{I}} = [0.9, 0]$, which does not satisfy $(\exists \mathtt{r}.\mathtt{A})^{\mathcal{I}} \leq \mathtt{B}^{\mathcal{I}}$, so hierarchical loss is 0.9. Through gradient decent, until loss becomes 0, $\mathtt{A}^{\mathcal{I}}$ is decreased as $[0, 0]$, and $\mathtt{B}^{\mathcal{I}}$ is increased as $[0.9, 0]$. $\mathtt{A}^{\mathcal{I}}$ is not expected to be changed and $\mathtt{B}^{\mathcal{I}}$ is expected to be increased as $[0.9, 0]$. In $\mathcal{O}_2$, $(\forall \mathtt{r}.\mathtt{A})^{\mathcal{I}} = [0.9, 1]$, $\mathtt{A}^{\mathcal{I}}$ will be revised as $[0, 0]$, $\mathtt{B}^{\mathcal{I}}$ will be revised as $[0.5, 1]$, and $\mathtt{r}^{\mathcal{I}}(s_1, s_2) = 1$. In $\mathcal{O}_3$ and $\mathcal{O}_4$, there is no revision.

Example 3. Given a perceptual grounding $\mathcal{I}$ in the domain $\{s_1, s_2\}$, $\mathtt{A}^{\mathcal{I}} = [0, 0.9]$ and $\mathtt{B}^{\mathcal{I}} = [0.9, 0]$, $\mathtt{r}^{\mathcal{I}}(s_1, s_2) = \mathtt{r}^{\mathcal{I}}(s_1, s_1) = \mathtt{r}^{\mathcal{I}}(s_2, s_1) = \mathtt{r}^{\mathcal{I}}(s_2, s_2) = 0$.

According to the semantics of fuzzy $\mathcal{ALC}$, $(\exists \mathtt{r}.\mathtt{A})^{\mathcal{I}} = (\forall \mathtt{r}.\mathtt{A})^{\mathcal{I}} = [0, 0]$, which satisfies $(\exists \mathtt{r}.\mathtt{A})^{\mathcal{I}} \leq \mathtt{B}^{\mathcal{I}}$, so no revision is executed. In $\mathcal{O}_3$, $\mathtt{B}^{\mathcal{I}}$ and $\mathtt{r}^{\mathcal{I}}$ is revised if there are other s_n that $\mathtt{A}^{\mathcal{I}}$ is not zero. In $\mathcal{O}_4$, there is no revision.

3 Experiment Setting

In the masked ABox revision task, we used 6 ontologies ("Ontodm" and "Nifdys" are not consistent), while in the conjunctive query answering task, we used 4 consistent ontologies.

The ontologies used for the experiments are taken from Bioportal[1], which, currently, includes more than 700 biomedical ontologies from different sources. We require the ontologies to have at least the logical operator of negation, disjunction, or universal quantifier, as well as 100 ABox assertions. Five ontologies fall into this set, with two of them ("Ontodm" and "Nifdys") not consistent in some assertions; it remains to see whether DF-$\mathcal{ALC}$ would revise these errors. A taxonomy ontology (Sso) is also added for comparison. We also test a terseness ontology "Family", which contains multiple instantiated families but its knowledge is incomplete. Based on "Family", we augment it into "Family2" by adding some knowledge that can bridge with the instantiation. The information about these ontologies is shown in Table 1. Adam optimizer was used with a learning rate of 2e-4 to learn the grounding. Early stopping with 10 epochs tolerance was used to limit the running time.

The mask rate of ABox ranges from $\{20\%, 40\%, 60\%, 80\%\}$. We set the unknown region as $[0.2, 0.8]$. Meanwhile, the truth values greater (less) than $\alpha = 0.8$ $(1 - \alpha = 0.2)$ were assumed to be true (false).

We generated 20 queries in each form, and the answer set of each query was not empty. Considering the time complexity of using a logical reasoner to get the true answer set, we only used two forms of conjunctive queries (CQs) in-depth 2 (the depth is determined by the conjunction amounts in the query). We chose all individuals with

[1] http://bioportal.bioontology.org/ontologies

	Family	Family2	GlycoRDF	Nifdys	Nihss	Ontodm	Sso
# TBox axioms	2032	2054	1453	6435	318	3476	2050
# ABox axioms	224	224	518	2920	146	1113	366
# Concepts	19	19	113	2751	18	838	176
# Roles	4	4	91	68	16	78	22
# Individuals	202	202	219	102	106	187	158
Expressivity	$\neg$	$\neg, \sqcap, \exists$	$\neg, \sqcup, \exists$	$\neg, \sqcup, \exists$	$\neg$	$\neg, \sqcap, \sqcup, \exists, \forall$	/

Table 1: Ontology information

$Q^{\mathcal{I}}(a) \geq 0.8$ to be the answer for query Q. And use the answers generated by logical reasoner as ideal answers to evaluate the predicted answers with precision and recall as metrics.

4 Related Works

4.1 Neuro-symbolic Computing

Neural-symbolic computing aims at computing with both learning and reasoning abilities, to step towards the combination of symbolic and sub-symbolic systems. Current learning ability relies largely on differentiable programming to draw conclusions from observations and apply them, while current reasoning ability relies largely on logical programming to give conclusions inferred from premises and rules through deductive reasoning, give rules according to observations comprising premises and conclusions through inductive reasoning, and give premises that can interpret conclusions according to rules through abductive reasoning. So it comes with challenges in the integration and representation of these two kinds of programming paradigms. From the perspective of integration, research works differ in logical techniques that are mainly consumed. Neural-symbolic inductive logical programming [Wang et al., 2013, Bühmann et al., 2016, Yang et al., 2017, Evans and Grefenstette, 2018, Sen et al., 2022] and statistical relational learning (e.g. Markov logic network [Richardson and Domingos, 2006], probabilistic soft logic [Bach et al., 2017]) works seek to learn probabilistic logical rules from observations. This requires learning model parameters in a continuous space and the structure in a discrete space. SATNet [Wang et al., 2019] learns rules from labeled data by transforming the learning problem as SAT problem[2]. To combine the ability of deductive reasoning, the first line of research learn to reason by modeling the inference procedure using neural networks or replacing logical computations with differentiable functions [Towell and Shavlik, 1994, Hölldobler et al., 1999, Rocktäschel and Riedel, 2016, 2017, Diligenti et al., 2017, Ebrahimi et al., 2021]. But this neglects factual knowledge which bridges the physical world and the conceptual world, so the second line of research aims to find an interpretation (grounding) that satisfies

[2] this is the satisfiability (SAT) problem which aims to determine whether there exists an interpretation that satisfies a given formula

theories which can be a mapping between these two worlds by encoding the satisfiability of theories in the loss function [Badreddine et al., 2022, Serafini and Garcez, 2016, Riegel et al., 2020, Topan et al., 2021, van Krieken et al., 2022]. The notable work Logical Tensor Network (LTN) [Badreddine et al., 2022] uses neural networks to represent the fuzzy function and predicates of theories, which is learned from labeled data. To solve the symbol grounding problem, LTN learns the interpretation with trained parameters that can maximize the satisfiability of theories. But these works cannot find explanations of observations according to theories, so abductive learning-based neural-symbolic works are proposed to use the explanations getting through abductive reasoning to promote the interpretability of the computing [Zhou, 2019, Huang et al., 2020, Tsamoura et al., 2021, Cai et al., 2021]. From the view of representation, some works are based on classical logic — propositional logic [Towell and Shavlik, 1994, Zhou, 2019, Tsamoura et al., 2021, Cai et al., 2021], description logic [Bühmann et al., 2016, Eberhart et al., 2019, Ebrahimi et al., 2021], or first-order logic [Hölldobler et al., 1999, Wang et al., 2013, Rocktäschel and Riedel, 2016, Serafini and Garcez, 2016, Rocktäschel and Riedel, 2017, Yang et al., 2017, Evans and Grefenstette, 2018, Sen et al., 2022], others are based on non-classical logic, such as fuzzy logic [Diligenti et al., 2017, Riegel et al., 2020, van Krieken et al., 2022], or probabilistic logic [Wang et al., 2013, Manhaeve et al., 2018]. Our work tries to combine the deductive reasoning ability in a novel way, which follows the work of LTN. As stated in the introduction, when using declarative knowledge, data bias should be ignored, so different from works that only maximize the satisfiability of the knowledge base, we give a compulsory way to inject knowledge.

Bibliography

Stephen H Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. Hinge-loss markov random fields and probabilistic soft logic. 2017.

Samy Badreddine, Artur d'Avila Garcez, Luciano Serafini, and Michael Spranger. Logic tensor networks. *Artificial Intelligence*, 303:103649, 2022.

Lorenz Bühmann, Jens Lehmann, and Patrick Westphal. Dl-learner—a framework for inductive learning on the semantic web. *Journal of Web Semantics*, 39:15–24, 2016.

Le-Wen Cai, Wang-Zhou Dai, Yu-Xuan Huang, Yu-Feng Li, Stephen Muggleton, and Yuan Jiang. Abductive learning with ground knowledge base. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1815–1821, 2021.

Michelangelo Diligenti, Marco Gori, and Claudio Sacca. Semantic-based regularization for learning and inference. *Artificial Intelligence*, 244:143–165, 2017.

Aaron Eberhart, Monireh Ebrahimi, Lu Zhou, Cogan Shimizu, and Pascal Hitzler. Completion reasoning emulation for the description logic el+. *arXiv preprint arXiv:1912.05063*, 2019.

Monireh Ebrahimi, Aaron Eberhart, and Pascal Hitzler. On the capabilities of pointer networks for deep deductive reasoning. *arXiv preprint arXiv:2106.09225*, 2021.

Richard Evans and Edward Grefenstette. Learning explanatory rules from noisy data. *Journal of Artificial Intelligence Research*, 61:1–64, 2018.

Steffen Hölldobler, Yvonne Kalinke, and Hans-Peter Störr. Approximating the semantics of logic programs by recurrent neural networks. *Applied Intelligence*, 11(1): 45–58, 1999.

Yu-Xuan Huang, Wang-Zhou Dai, Jian Yang, Le-Wen Cai, Shaofen Cheng, Ruizhang Huang, Yu-Feng Li, and Zhi-Hua Zhou. Semi-supervised abductive learning and its application to theft judicial sentencing. In *2020 IEEE international conference on data mining (ICDM)*, pages 1070–1075. IEEE, 2020.

Robin Manhaeve, Sebastijan Dumancic, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. Deepproblog: Neural probabilistic logic programming. *Advances in Neural Information Processing Systems*, 31, 2018.

Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine learning*, 62(1):107–136, 2006.

Ryan Riegel, Alexander Gray, Francois Luus, Naweed Khan, Ndivhuwo Makondo, Ismail Yunus Akhalwaya, Haifeng Qian, Ronald Fagin, Francisco Barahona, Udit Sharma, et al. Logical neural netwroks. *arXiv preprint arXiv:2006.13155*, 2020.

Tim Rocktäschel and Sebastian Riedel. Learning knowledge base inference with neural theorem provers. In *Proceedings of the 5th workshop on automated knowledge base construction*, pages 45–50, 2016.

Tim Rocktäschel and Sebastian Riedel. End-to-end differentiable proving. *Advances in neural information processing systems*, 30, 2017.

Prithviraj Sen, Breno WSR de Carvalho, Ryan Riegel, and Alexander Gray. Neuro-symbolic inductive logic programming with logical neural networks. In *Proceed-*

ings of the AAAI Conference on Artificial Intelligence, volume 36, pages 8212–8219, 2022.

Luciano Serafini and Artur d'Avila Garcez. Logic tensor networks: Deep learning and logical reasoning from data and knowledge. *arXiv preprint arXiv:1606.04422*, 2016.

Sever Topan, David Rolnick, and Xujie Si. Techniques for symbol grounding with satnet. *Advances in Neural Information Processing Systems*, 34:20733–20744, 2021.

Geoffrey G Towell and Jude W Shavlik. Knowledge-based artificial neural networks. *Artificial intelligence*, 70(1-2):119–165, 1994.

Efthymia Tsamoura, Timothy Hospedales, and Loizos Michael. Neural-symbolic integration: A compositional perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 5051–5060, 2021.

Emile van Krieken, Erman Acar, and Frank van Harmelen. Analyzing differentiable fuzzy logic operators. *Artificial Intelligence*, 302:103602, 2022.

Po-Wei Wang, Priya Donti, Bryan Wilder, and Zico Kolter. Satnet: Bridging deep learning and logical reasoning using a differentiable satisfiability solver. In *International Conference on Machine Learning*, pages 6545–6554. PMLR, 2019.

William Yang Wang, Kathryn Mazaitis, and William W Cohen. Programming with personalized pagerank: a locally groundable first-order probabilistic logic. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 2129–2138, 2013.

Fan Yang, Zhilin Yang, and William W Cohen. Differentiable learning of logical rules for knowledge base reasoning. *Advances in neural information processing systems*, 30, 2017.

Zhi-Hua Zhou. Abductive learning: towards bridging machine learning and logical reasoning. *Science China Information Sciences*, 62(7):1–3, 2019.