Ambiguity-Aware Abductive Learning

Hao-Yuan He, Hui Sun, Zheng Xie, Ming Li

ICML International Conference On Machine Learning

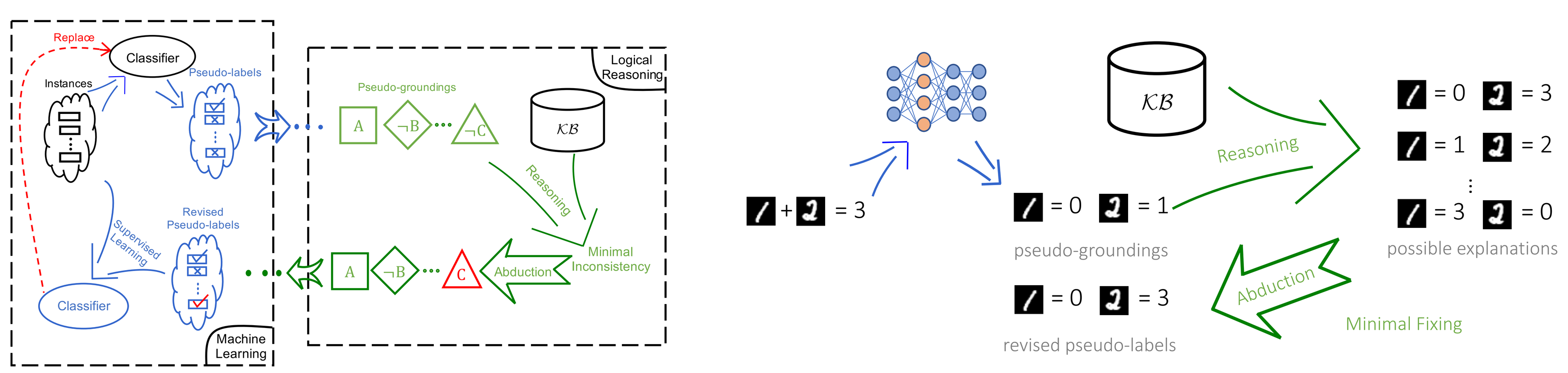
LAMDA Learning And Mining from DatA
http://www.lamda.nju.edu.cn

Ambiguity in Abductive Learning

Abductive learning (ABL) combining machine learning and logical reasoning mutually beneficial [1].

- *Full perception ability*: the perception model can be any form, e.g., neural networks (NNs).
- *Full logical reasoning ability*: it can use any form of knowledge or rules, e.g., first order logic (FOL).

Due to the above properties, ABL has been successfully applied in various practical tasks, including theft judicial sentencing [2], stroke evaluation [3], optical character recognition [4], and historical document segmentation and recognition [5].



(a) Framework of ABL [1] (b) Ambiguity: Abduction process yields various valid explanations.

Toy Example of ABL. The digit equation $\texttt{SUM}(1, 2) = 3$ is presented, with the background knowledge that it represents an addition task. Initially, the machine learning model identifies $1 = 1$ and $2 = 1$. However, it is deduced that $1 + 1 \neq 3$. Nonetheless, a plausible explanation is proposed: $1 = 1, 2 = 2$, aligning with the background knowledge and validating the equation.

Ambiguity. The term *ambiguity* in this context means that the abduction process yields not only the correct result but also other validate candidates [6]. Take the example above, the validate candidates $[1 = 0, 2 = 3], \cdots, [1 = 3, 2 = 0]$, which are also reasonable hypotheses.

Challenge from Ambiguity. Our model may be trapped in an shortcut, stems from the ambiguity of the abduced result: once the model selects an incorrect candidate, the error can be strengthened in the subsequent process.

Problem Formulation

The perception model $f : \mathcal{X}^m \mapsto \mathcal{Z}^m$, maps from the input space to the symbol space; The knowledge base KB, is comprised of rules defined over the symbol space; Indirect information $y \in \mathcal{Y}$, the target label, such that $\boldsymbol{z} \wedge \text{KB} \models y$. Prior ABL methods basically solve an ERM problem:

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(f(\boldsymbol{x}^{(i)}), \boldsymbol{z}^{(i)}\right)$$
$$\text{s.t.} \quad \boldsymbol{z}^{(i)} = \arg\min_{\boldsymbol{c} \in \boldsymbol{s}^{(i)}} \text{Score}\left(\boldsymbol{c}, f(\boldsymbol{x}^{(i)})\right), i \in [N], \tag{1}$$

where $\boldsymbol{s}^{(i)} = \{\boldsymbol{c} | \boldsymbol{c} \wedge \text{KB} \models y^{(i)}\}$. The term $\text{Score}\left(\boldsymbol{c}, f(\boldsymbol{x}^{(i)})\right)$ is used to select the most likely correct candidate $\boldsymbol{c}$ based on the model's prediction $f(\boldsymbol{x}^{(i)})$. For instance, [7] use the Hamming distance as the score function:

$$\text{Score}\left(\boldsymbol{c}, f(\boldsymbol{x}^{(i)})\right) = \text{Hamming}\left(\boldsymbol{c}, f(\boldsymbol{x}^{(i)})\right). \tag{2}$$

This function tries to use the candidate $\boldsymbol{c}^{*}$ that has most of the same labels as the prediction. Also, it is feasible to extend this approach to include the confidence provided by the model's prediction as follows:

$$\text{Score}(\boldsymbol{c}, f(\boldsymbol{x}^{(i)})) = 1 - \prod_{j=1}^{m} f(\boldsymbol{x}^{(i)})_{c_j}. \tag{3}$$

In this context, $f(\cdot)$ represents the model's predicted distribution over $\mathcal{Z}$. Additionally, $f(\boldsymbol{x}^{(i)})_{c_j}$ is the model's estimated probability of the class c_j given the input $x_j^{(i)}$.

Contact: hehy@lamda.nju.edu.cn

Proposed Method

Challenge. Prior works overlook the ambiguity in the abduction process and is prone to error when it fails to identify the correct candidates.

Main Idea. Rather than considering one candidate at once, we use all valid candidates.

Ambiguity-Aware Abductive Learning (A^{3}BL). Our goal is to estimate the class probability distribution $p(z|x, \boldsymbol{s}) = [p(z = 1|x, \boldsymbol{s}), \cdots, p(z = L|x, \boldsymbol{s})]^{\top}$ for training the perception model from the ambiguous abduction results $\boldsymbol{s} = \{\boldsymbol{c}_i | \boldsymbol{c}_i \wedge \text{KB} \models y\}$. The probability of a candidate label sequence can be formalized as:

$$p(\boldsymbol{z} = \boldsymbol{c}_i | \boldsymbol{x}) = \prod_{1 \leq j \leq m} p(z_j = c_{ij} | x_j). \tag{4}$$

The posterior probability of a candidate label sequence $\boldsymbol{c}_i$, given the candidate set $\boldsymbol{s}$, can be obtained by:

$$p(\boldsymbol{z} = \boldsymbol{c}_i | \boldsymbol{x}, \boldsymbol{s}) = \frac{p(\boldsymbol{z} = \boldsymbol{c}_i | \boldsymbol{x})}{\sum_{\boldsymbol{c} \in \boldsymbol{s}} p(\boldsymbol{z} = \boldsymbol{c} | \boldsymbol{x})}. \tag{5}$$

For an instance x_j appeared in the sequence $\boldsymbol{x}$, its probability of being k-th class should be:

$$p(z = k | x_j, \boldsymbol{s}) = \sum_{\boldsymbol{c}_i \in \boldsymbol{s}} \mathbb{I}[k = c_{ij}] p(\boldsymbol{z} = \boldsymbol{c}_i | \boldsymbol{x}, \boldsymbol{s}). \tag{6}$$

Minimize the difference between the model's output and the probability distribution:

$$\frac{1}{m \cdot N} \sum_{i=1}^{N} \sum_{j=1}^{m} \mathcal{L}_{\text{cls}}\left(f(x_j^{(i)}), p(z | x_j^{(i)}, \boldsymbol{s}^{(i)})\right). \tag{7}$$

The above object, named empirical ambiguity-aware abductive risk, denotes as $\hat{R}_{\text{A}^3}$. By optimizing this object, A^{3}BL can fully leverage the ambiguous abductive results to facilitate learning.

Require: knowledge base KB, perception model f, dataset $(\hat{\mathcal{X}}^m, \hat{\mathcal{Y}})$.
for each pair $(\boldsymbol{x}^{(i)}, y^{i})$ in $(\hat{\mathcal{X}}^m, \hat{\mathcal{Y}})$ **do**
  $\tilde{\boldsymbol{z}}^{(i)} \leftarrow f(\boldsymbol{x}^{(i)})$
  $\tilde{y}^{(i)} \leftarrow$ KB.logical_forward$(\tilde{\boldsymbol{z}}^{(i)})$
  if $y^{(i)} \neq \tilde{y}^{(i)}$ **then**
    $\boldsymbol{s}^{(i)} \leftarrow$ KB.abduce$(y^{(i)}, \tilde{\boldsymbol{z}}^{i})$ ▷ Abduce all candidates that valid.
  end if
  Estimate class probability distribution $p(z|x, \boldsymbol{s})$ based on *all* candidates.
  loss $\leftarrow \frac{1}{m \cdot N} \sum_{i=1}^{N} \sum_{j=1}^{m} \mathcal{L}_{\text{cls}}\left(f(x_j^{(i)}), p(z | x_j^{(i)}, \boldsymbol{s}^{(i)})\right)$ ▷ Ambiguity-aware abductive risk.
  Update the perception model f
end for

Algorithm: Overall Algorithm of A^{3}BL

Guarantee: Learning from Ambiguous Abduction is Possible. *Suppose that $\mathcal{L}(f(\boldsymbol{x}), \boldsymbol{z})$ is ρ-Lipschitz with respect to $f(\boldsymbol{x})$ for all $\boldsymbol{z} \in \mathcal{Z}^m$ and upper-bounded by $M = \sup_{f \in \mathcal{F}, \boldsymbol{x} \in \mathcal{X}^m, \boldsymbol{z} \in \mathcal{Z}^m} \mathcal{L}(f(\boldsymbol{x}), \boldsymbol{z})$. Let $\mathfrak{R}_n(\mathcal{F}_i)$ be the Rademacher complexity of $\mathcal{F}_i$ with sample size n. Then for any $\delta > 0$, with probability at least $1 - \delta$,*

$$R(\hat{f}_{\text{A}^3}) - R(f^{\star}) \leq 4\sqrt{2}\rho \sum_{i=1}^{K} \mathfrak{R}_n(\mathcal{F}_i) + 2M\sqrt{\frac{\log(2/\delta)}{2n}}. \tag{8}$$

As $n \to \infty$, it follows that $\mathfrak{R}_n(\mathcal{F}_i) \to 0$ for all parametric models with a bounded norm, such as deep networks trained with weight decay [8]. Therefore,the above theorem indicates that the learning process is consistent in an asymptotic sense.

Experiments

Remark 1. ABL suffers from ambiguity, and thus it is easy to fall into suboptimal shortcuts, especially when training from scratch.

Remark 2. A^{3}BL beats other methods in terms of *stability* and *convergence speed*.

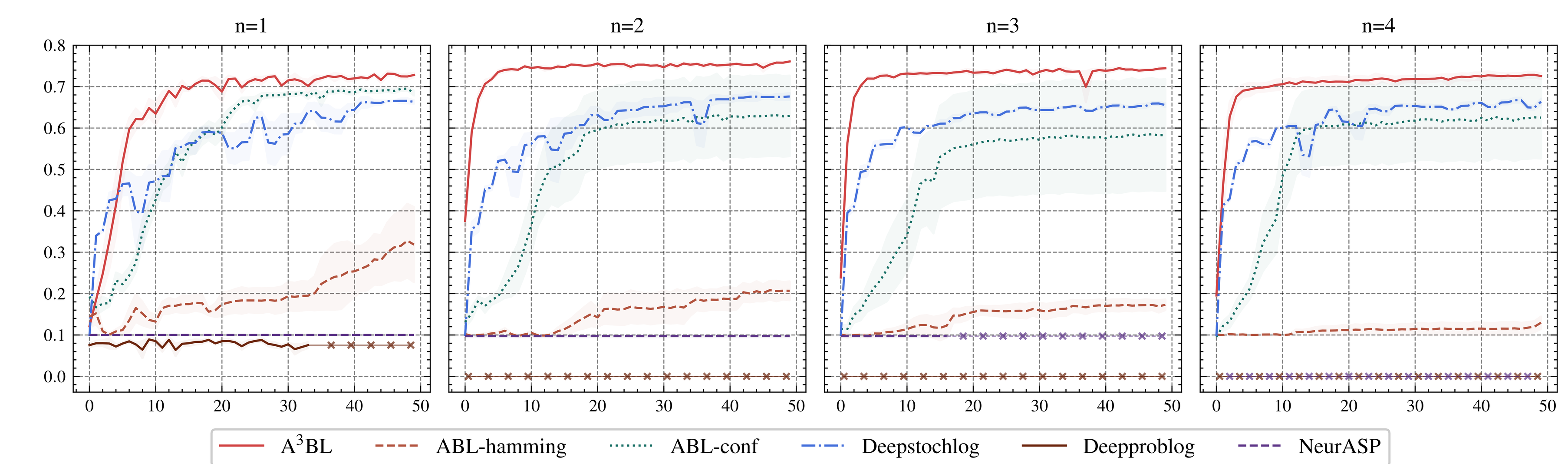


Figure: *Performance curve of the Addition task.* A^{3}BL (red line) performs better.

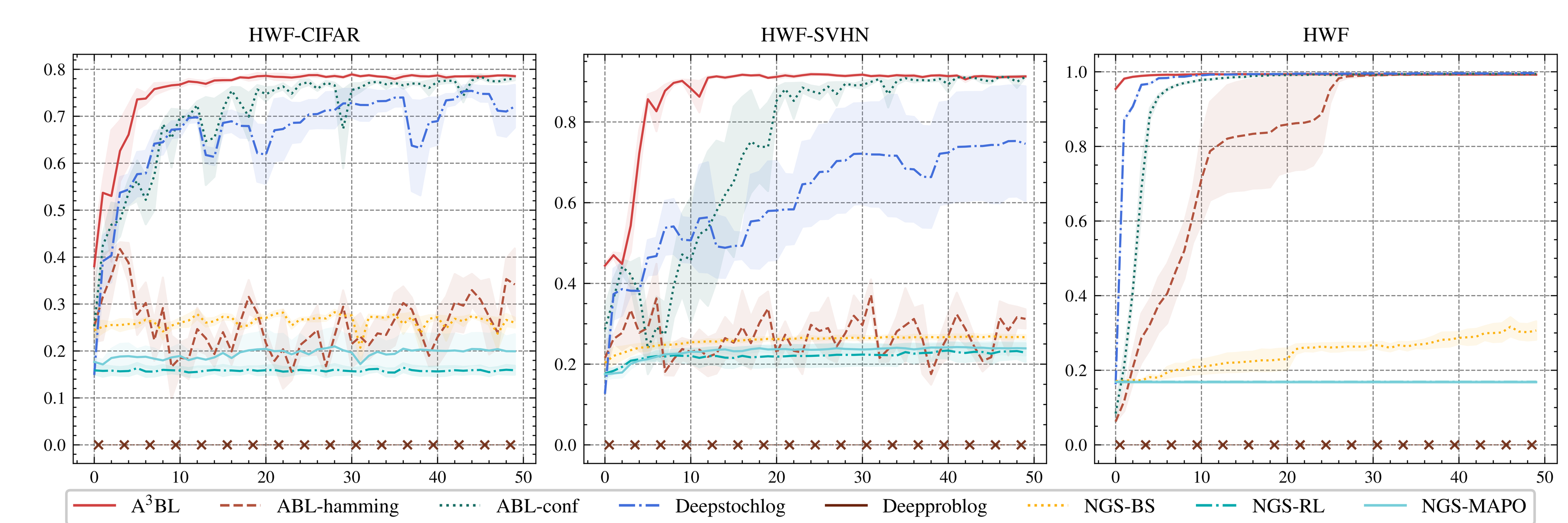
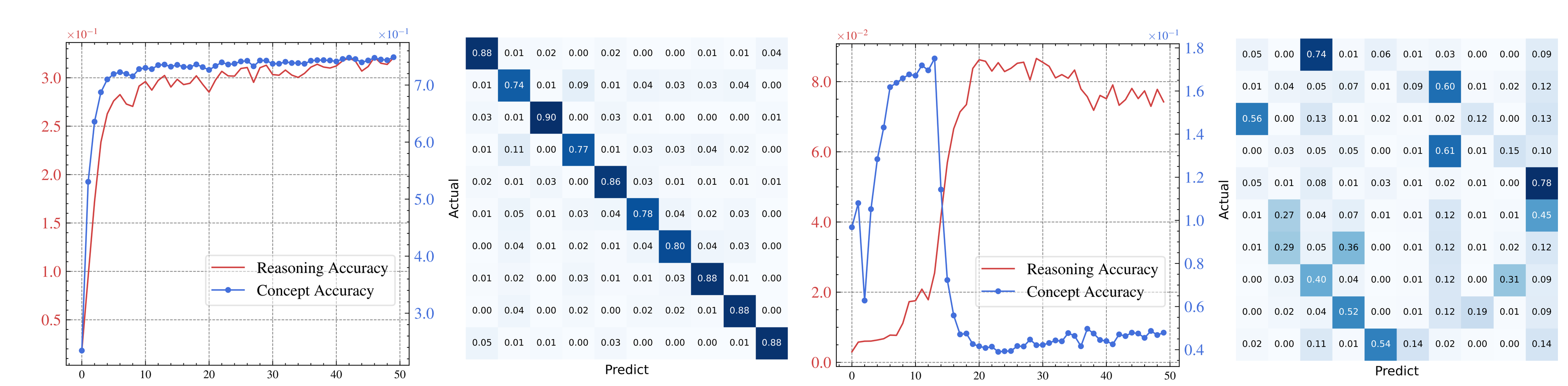


Figure: *Performance curve of the HWF task.* A^{3}BL (red line) performs better.



(a) A^{3}BL (b) ABL-conf

Figure: ABL fails in reasoning shortcuts, while A^{3}BL avoids it.

Take Home Messages

- The abductive learning framework faces challenge of ambiguity during the abduction process.
- We first point out and address this problem by considering all possible candidates.
- Theoretical guarantee: learning from ambiguous abduction results is possible.
- A^{3}BL demonstrates faster convergence, superior stability, and better performance.

[1] Zhi-Hua Zhou. Abductive Learning: Towards Bridging Machine Learning and Logical Reasoning. *Science China Information Sciences*, 62(7):76101, 2019.
[2] Yu-Xuan Huang et al. Semi-supervised Abductive Learning and Its Application to Theft Judicial Sentencing. In *Proceedings of the 20th IEEE International Conference on Data Mining*, pages 1070–1075, Sorrento, Italy, 2020.
[3] Jiachen Wang et al. Tac-Valuer: Knowledge-based Stroke Evaluation in Table Tennis. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 3688–3696, Virtual Event, Singapore, 2021.
[4] Le-Wen Cai et al. Abductive Learning with Ground Knowledge Base. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, pages 1815–1821, Virtual, 2021.
[5] En-Hao Gao et al. Knowledge-enhanced Historical Document Segmentation and Recognition. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, pages 8409–8416, Vancouver, BC, Canada, 2024.
[6] Lorenzo Magnani. *Abductive Cognition - The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*, volume 3 of *Cognitive Systems Monographs*. Springer, Berlin, Heidelberg, 2009.
[7] Wang-Zhou Dai et al. Bridging Machine Learning and Logical Reasoning by Abductive Learning. In *Advances in Neural Information Processing Systems 32*, pages 2815–2826, Vancouver, BC, Canada, 2019.
[8] Nan Lu et al. On the Minimal Supervision for Training Any Binary Classifier from Only Unlabeled Data. In *Proceedings of 7th International Conference on Learning Representations*, New Orleans, LA, USA, 2019.