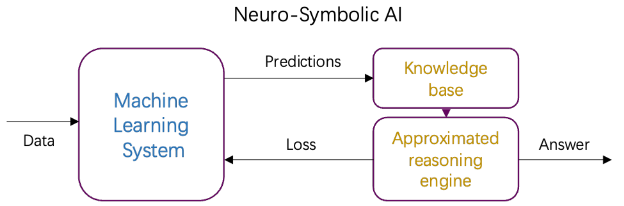
Reduced Implication-bias Logic Loss for Neuro-Symbolic Learning

Hao-Yuan He and **Wang-Zhou Dai** and **Ming Li**
Nanjing University, Nanjing, 210023, China

Logic Rules as Loss Functions

Neuro-Symbolic (NeSy) AI aims to bridge the gap between neural networks and symbolic reasoning to achieve a more comprehensive form of artificial intelligence.
Researchers have proposed approximating logical reasoning with differentiable operators (e.g., fuzzy operators) in order to transform symbolic knowledge into loss functions. Models can then be trained using end-to-end.



Fuzzy based Logic Loss

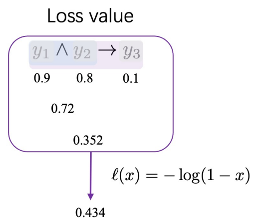
Fuzzy logic is a form of many-valued logic in which the truth value of variables may be any real number between 0 and 1. It replace exact logical operator with approximate differentiable operator. The commonly used fuzzy operators in the field of NeSy is Reichenbach. Below is an example.

Model Prediction y_1, y_2, y_3

Rule $y_1 \wedge y_2 \rightarrow y_3$

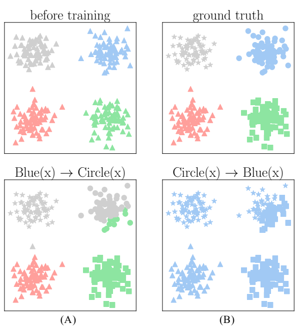
Function	Expression
$\text{Not}(p)$	$1-p$
$\text{And}(p,q)$	$p \cdot q$
$\text{Or}(p,q)$	$p+q-p\cdot q$
$\text{Imply}(p,q)$	$1-p+p\cdot q$

Reichenbach operators

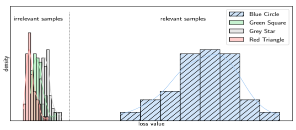


Optimization objective $R_{\text{task}}(f) + \lambda \cdot R_{\text{logic}}(f)$

Implication Bias: Case Study



- Setting:
  Each instance has two attributes color and shape. Training process only use the shape labels.
  Case A,B are trained with different logic loss (rules).
- Findings:
  Before training, color-labels are correctly predict.
  After training, shape-labels are correctly predict but the premises of the implication rules are unsatisfied.
- Why:
  The model can satisfy the logical constraints easily through just negate the premises of the implication rules without need to consider consequents of the implication rules.

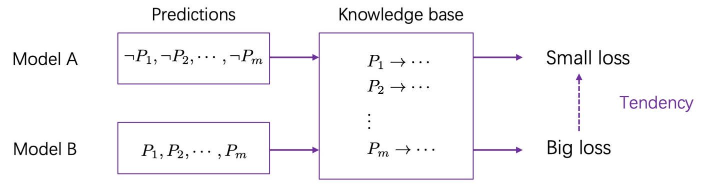


Distribution of different samples given by the logic loss values. Computed from case B

- Findings:
  small loss values samples are irrelevant samples

Implication Bias

The implication bias is a tendency for NeSy systems to negate the premises of implication rules in order to increase consistency with the knowledge base.

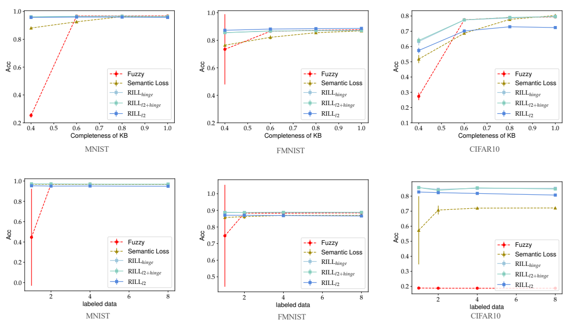


(Implication Biased). A logic loss is said to be implication biased if there exists a small $\delta > 0$ such that its implication likelihood $I(x, y)$ satisfies $\forall x \in (0, \delta), \exists y \in [0, 1]$,

$$\frac{\partial I}{\partial x} < 0$$

Unofficially, this means that in almost all cases there is a tendency for the logic loss to negate the truth value of the premises in order to lowering the loss.

Empirical Study



RILL is more stable and performs better than other methods, particularly when the knowledge base is incomplete or supervised data is limited.

Reduced Implication-bias Logic Loss

Insight.
Samples with low confidence in their premise should be assigned less importance.

- These samples tend to have lower loss values, which means they contribute a little to the optimization objective compared with other samples that have higher loss values.
- Most of these samples are irrelevant samples corresponding to this rule, which means they are less important compared with other relevant samples.

low confidence premise samples + rule satisfied samples = small loss samples

Rank loss values from low to high

$\ell_1^r, \ell_2^r, \cdots, \ell_t^r, \cdots, \ell_N^r.$

assign less importance

Hinge $\text{Average}(\{\ell_i^r \mid \ell_i^r \geq \epsilon, i \in [N]\})$

L2 $\text{Average}(\{(\ell_i^r)^2 \mid i \in [N]\})$

Hinge + L2 $\text{Average}(\{\mathcal{T}(\ell_i^r) \mid i \in [N]\})$

$$\mathcal{T}(\ell) = \ell^2 \cdot \mathbb{I}[\ell \leq \epsilon] + \ell \cdot \mathbb{I}[\ell > \epsilon].$$

Take Home Message

- Approximating logical reasoning using fuzzy based operators can bring implication bias.
- Reduced Implication-bias Logic Loss (RILL for short) to mitigate this bias through assign less importance to those samples have small logic loss values.
- RILL performs better on both incomplete knowledge base and insufficient supervision cases.