Bridging Machine Learning and Logical Reasoning by Abductive Learning

Wang-Zhou Dai†, Xu Qiuling†, Yang Yu† and Zhi-Hua Zhou

National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

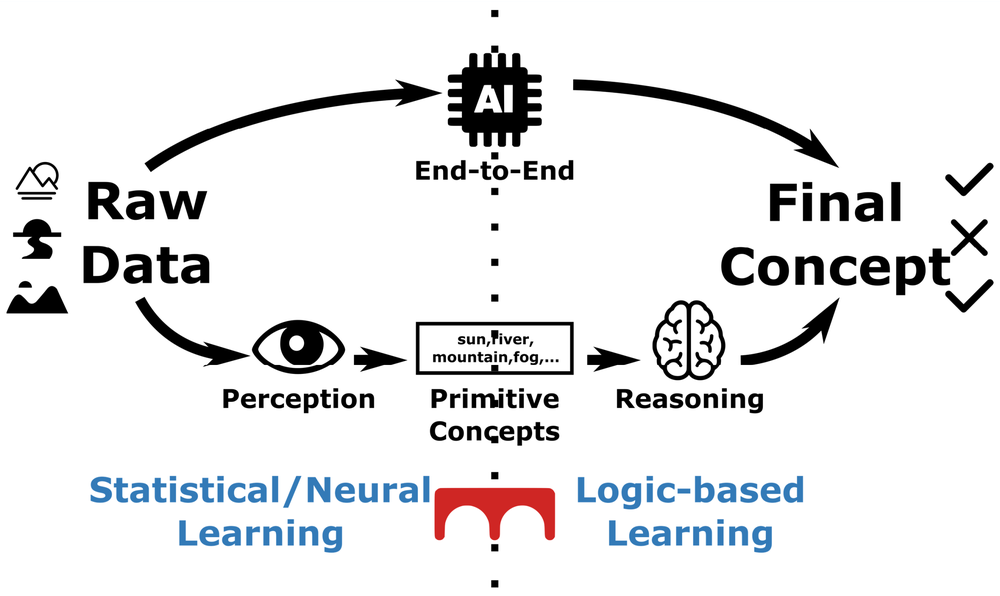
I. Motivation

Perception and **reasoning** are two abilities of intelligence that are integrated seamlessly. In AI, perception is usually **statistics/neural-based learning** and reasoning is often formalized by **logic-based AI**. However, it is difficult to combine these two learning paradigms:

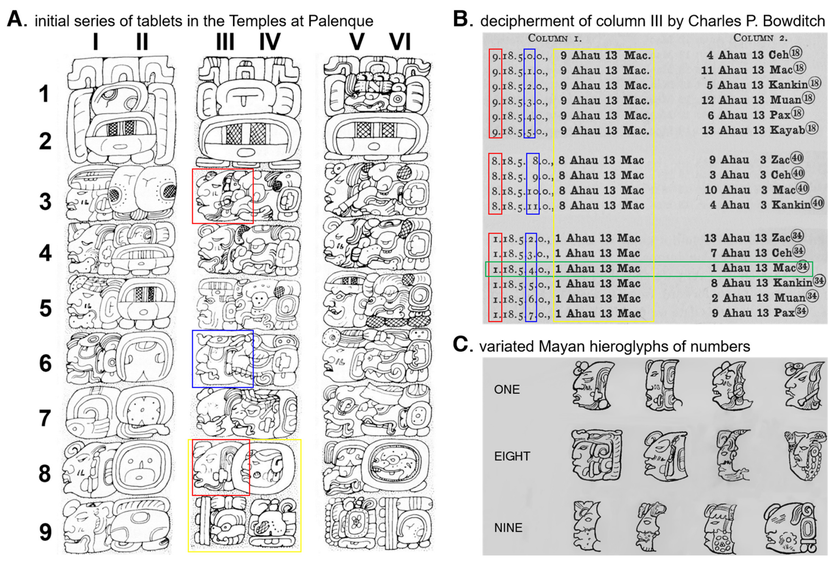
1. Statistical/Neural learning are difficult to incorporate general **background knowledge**;
2. Probabilistic logic-based learning is hard to handle **raw data** and **noisy domain**;
3. Neural symbolic learning designs **gradient operators** to propagate training errors during the approximate ***"soft logical reasoning"***, which: a) loses the power of **extrapolation**, b) introduces **independent assumptions**, and c) requires **pre-compiled** network structures.

The **abductive learning (ABL)** framework targeted at unifying the two learning paradigms in a mutual beneficial way, it has 3 major components:

1. **Machine learning models** learn to perceive primitive concepts from raw data;
2. **Logical abduction** handles background knowledge, correct the wrongly perceived facts for improving the machine learning models, and learns relational models;
3. **Consistency optimization** that maximizes the consistency of machine learning models, abduced results and background knowledge.

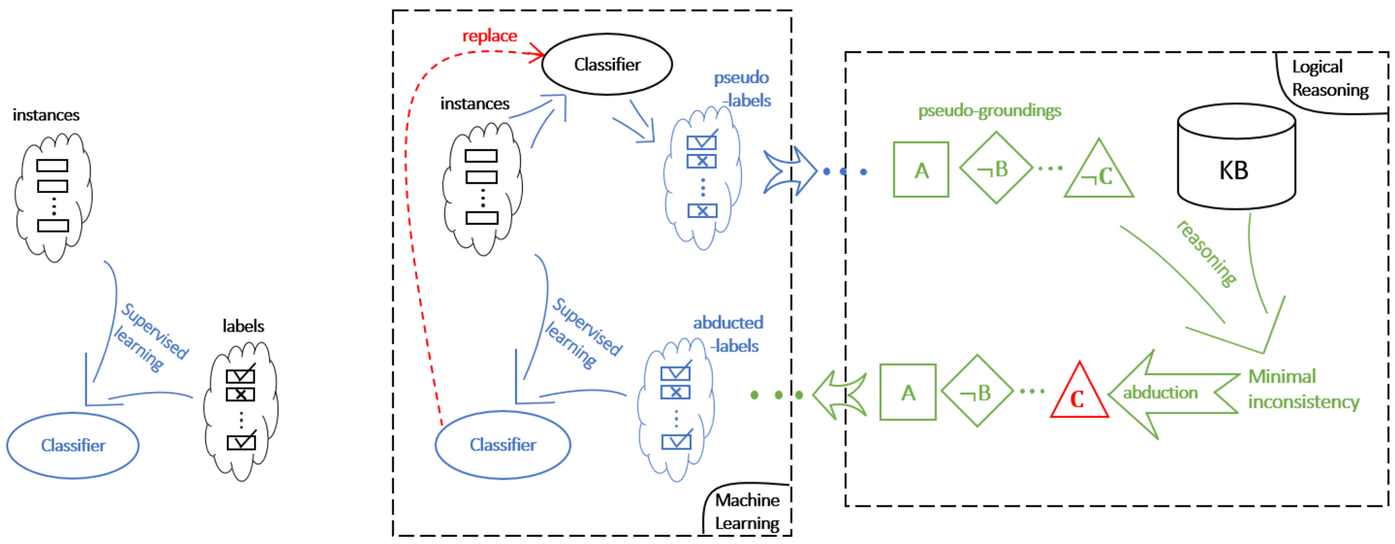


II. Human Abductive Problem-Solving



The Mayan tablet above are different arithmetic equations with unknown glyphs (now known as "head variants" of numbers), the figures show how Charles P. Bowditch deciphered the glyph. **(A)** The odd columns represent numbers and the even columns represent calendar time units. Rows 1-2 are initials, rows 3-6 represent the long count (a Mayan calendar system) of a time span, and rows 8-9 are the results computed from the initials and the time spans. **(B)** Conjectures and consistency checks made my Bowditch. Initially, he conjectured that III3 is 9, which resulted in inconsistency. Then he substituted those positions with other numbers and confirmed that the interpretation "1.18.5.4.0 1 Ahau 13 Mac" (in the green box) should be correct. **(C)** The highly varied character representations and unusual calendar system cause the decipherment of Mayan hieroglyphs to require both sensitive vision and a logical mind.

III. Framework



(a)Conventional supervised learning

Task: The input of **Abductive Learning** consists of training data $D = \{\langle \boldsymbol{x}_1, y_1 \rangle, \dots, \langle \boldsymbol{x}_n, y_n \rangle\}$ about a target concept C and a logical domain knowledgebase B, where $\boldsymbol{x}_i \in \mathcal{X}$ are raw data. The target concept C describes a certain relationship between a set of primitive concepts symbols $\mathcal{P} = \{p_1, \dots, p_r\}$ defined in B. The target is to output a hypothesis $H = p \cup \Delta_C$:

- $p: \mathcal{X} \mapsto \mathcal{P}$ is a mapping from the feature space to primitive symbols, i.e., it is a perception model formulated as a conventional machine learning model;
- Δ_C is a set of first-order logical clauses that defines the target concept C with $B \cup p(\boldsymbol{x}_i)$, which is called knowledge model.

(b) Abudctive learning

Challenge: No **direct label information** for learning p. To abduce the labels for training p, the correct Δ_C is required; To learn Δ_C for deducing the correct y, the accurate primitive facts $p(\boldsymbol{x})$ from p is required. **Difficult to train $p \cup \Delta_C$ simultaneously.**

Learning: Given a training example associated with a final output, logical abduction can abduce the possible primitive facts (**pseudo-labels**) based on background knowledge and the final output for training the machine learning model; And the consistent optimization is used for **maximizing the consistency** between the learned model and training data with regard to the background knowledge:

$$\max_{H=p\cup\Delta_C} \mathrm{Con}(H \cup D)$$

IV. Optimization

We define $\mathrm{Con}(H \cup D)$ as the size of subset $\widehat{D}_C \subseteq D$ which is consist with $H = p \cup \Delta_C$.

$$\widehat{D}_C = \max_{D_C \subseteq D} |D_C|$$
$$\text{s.t.} \quad \forall \langle \boldsymbol{x}_i, y_i \rangle \in D_C \quad (B \cup \Delta_C \cup p(\boldsymbol{x}_i) \models y_i).$$

However, when perception model p^t in the t-th epoch is undertrained, the **pseudo-labels** $p^t(\boldsymbol{X})$ could be incorrect and result in inconsistency. So some pseudo-labels

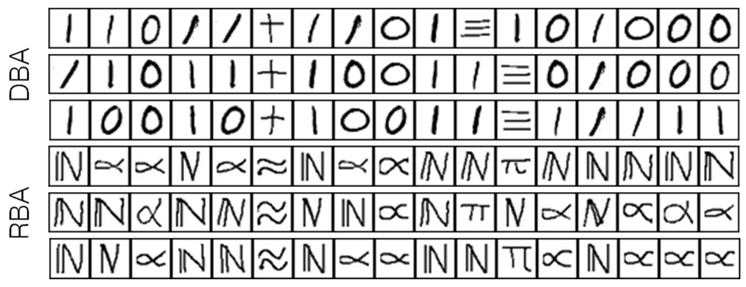
$\delta(p^t(\boldsymbol{X})) \subseteq p^t(\boldsymbol{X})$ should be **revised**, where δ is a function to estimate which pseudo-labels are wrong.

When the function marks the incorrect pseudo-labels, logical abduction can **abduce the revised pseudo-labels with a consistent** Δ_C^t. Therefore, we transform the objective into an optimization problem of **the heuristic error-guessing function** δ:

$$\max_{\delta} Con(p^t \cup \Delta_C(B, \delta(p(X)), Y) \cup D),$$
$$\text{s.t.} \quad |\delta(p^t(X))| \le M$$

The constant M stands for the maximum revision size of δ, it is used ensuring that the revision **should not be too far away from the perception** $p^t(X)$. The objective is non-convex, therefore we use **derivative-free optimization** approach to solve it.

V. Datasets



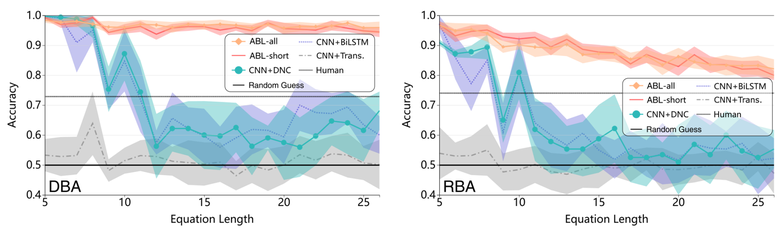
Each instance is a sequence of images forming a binary equation, associated with a label about the equation's correctness. **Labels of images** (i.e. the numeric symbols) are **NOT** available. The **meaning of "+"** (could be arithmetic plus or logical exclusive-or) is also **NOT** included in .

Length of equations are 5-26. We (**NLM-short**) only used 5-8.

Similar to the Mayan glyph decipherment task, we provide two kinds of background knowledge to Abductive Learning:

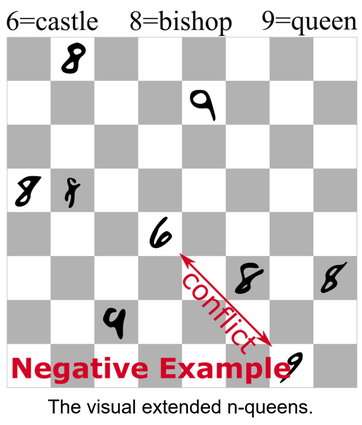
1. All equations have the form of $X + Y = Z$;
2. The "+" operator should be calculated bit-by-bit reversely.
3. Operation rules are **UNKNOWN**, i.e. results of 0+0, 0+1 and 1+0 could be 0, 1, 00, 01 or 10.

VI. Results



Result: Above results show that our approach have better generalization ability comparing to other state-of-the-art neural nets. Moreover, smaller training data (using only 5-8 length equations) does not degenerate the performance.

VII. Plug in classic symbolic AI tools



The visual extended n-queens.

ABL-ALP	ABL-CLP(FD)	ABL-CHR
97.6%	98.1%	97.9%
CNN	**Bi-LSTM**	
73.6%	51.3%	

ABL can take **classic symbolic AI systems** as reasoning module, e.g. **Constraint Logic Programs**.

In case that the main presenter (W.-Z. Dai) does not get his visa to attend this conference, please feel free to contact him via w.dai@imperial.ac.uk should you have any questions. Slides are available at http://daiwz.net →