Learning as Interpretation: Human vs Statistical Learning

Stephen Muggleton,
Imperial College London
### Human vs Statistical Learning

**UK EPSRC Priority 2016-2021 - Human-like Computing**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Human</th>
<th>Statistical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples per concept</td>
<td>Few ($\approx 1$)</td>
<td>Many ($\geq 10K$)</td>
</tr>
<tr>
<td></td>
<td>[Tenenbaum, 2011]</td>
<td></td>
</tr>
<tr>
<td>Concepts</td>
<td>Many ($\geq 10K$)</td>
<td>Few ($\approx 1$)</td>
</tr>
<tr>
<td></td>
<td>[Brown et al, 2008]</td>
<td></td>
</tr>
<tr>
<td>Background knowledge</td>
<td>Large</td>
<td>Small</td>
</tr>
<tr>
<td></td>
<td>[Brown, 2000]</td>
<td></td>
</tr>
<tr>
<td>Structure</td>
<td>Modular, re-useable</td>
<td>Monolithic</td>
</tr>
<tr>
<td></td>
<td>[Omrod et al, 2004]</td>
<td></td>
</tr>
</tbody>
</table>
Example 1: Dance Routine

<table>
<thead>
<tr>
<th>Observe</th>
<th>Perform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual perception</td>
<td>Motor program</td>
</tr>
</tbody>
</table>

- A girl watches a dance routine on television.
- Afterwards she reproduces the routine.
- The new dance moves are incorporated into her repertoire.
- Subsequent improvisation allows re-use of parts of routines.
Example 2: Learning words in a language

- Average undergraduate knows 20K words.
- Learning rate $= \frac{20000}{20 \times 365} = 2.7$ new words per day since birth.
- Presentations new word before assimilation $\approx 1$ [Zipf’s Law].
- Word assimilation involves visual, auditory, sense and context recognition of associated concept.
Learning as Interpretation

Write-once, Non-deterministic **Universal Turing Machine**

Computation = Learning = Interpretation = Perception
Meta-Interpretive Learning [IJCAI 2013]

**Prolog Meta-Interpreter** implements Learning as Interpretation.

**Input to Meta-Interpreter:** 1) Observations, 2) Meta-Rules, 3) Background Knowledge assignments (substitutions).

**Output from Meta-Interpreter:** Hypothesised assignments.

**Metagol supports** Problem decomposition by Predicate Invention and Learning recursion [MLJ 2015], Single example multi-task learning [ECAI 2014], Program Induction with resource and time-complexity optimisation [IJCAI 2015].
Generalised Meta-Interpreter

\[ \text{prove}([], BK, BK). \]
\[ \text{prove}([\text{Atom} | \text{As}], BK, BK_H) : - \]
\[ \text{metarule}(\text{Name}, \text{MetaSub}, (\text{Atom} \rightarrow \text{Body}), \text{Order}), \]
\[ \text{Order}, \]
\[ \text{save subst(metasub(\text{Name}, \text{MetaSub}), BK, BK_C)}, \]
\[ \text{prove}(\text{Body}, BK_C, BK_Cs), \]
\[ \text{prove}(\text{As}, BK_Cs, BK_H). \]
# Metarules

<table>
<thead>
<tr>
<th>Name</th>
<th>Meta-Rule</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instance</td>
<td>$P(X, Y) \leftarrow$</td>
<td>True</td>
</tr>
<tr>
<td>Base</td>
<td>$P(x, y) \leftarrow Q(x, y)$</td>
<td>$P \succ Q$</td>
</tr>
<tr>
<td>Chain</td>
<td>$P(x, y) \leftarrow Q(x, z), R(z, y)$</td>
<td>$P \succ Q, P \succ R$</td>
</tr>
<tr>
<td>TailRec</td>
<td>$P(x, y) \leftarrow Q(x, z), P(z, y)$</td>
<td>$P \succ Q$, $x \succ z \succ y$</td>
</tr>
</tbody>
</table>
Expressivity of $H_2^2$

Given an infinite signature $H_2^2$ has Universal Turing Machine expressivity [Tarnlund, 1977].

\[
\begin{align*}
\text{utm}(S,S) & \leftarrow \text{halt}(S). \\
\text{utm}(S,T) & \leftarrow \text{execute}(S,S_1), \text{utm}(S_1,T). \\
\text{execute}(S,T) & \leftarrow \text{instruction}(S,F), F(S,T).
\end{align*}
\]

Q: How can we limit $H_2^2$ to avoid the halting problem?
**Experimental applications**

**Vision**  Staircase [ILP 2013], Geometric Shape Learner [ILP 2015].

**Robotics**  Building stable walls [IJCAI 2013], Robot delivery and sorting [IJCAI 2013].

**Language**  Formal grammars [MLJ 2014], String transformations [ECAI 2014], Learning semantics [ILP 2015].
What next for Meta-Interpretive Learning?

Problem decomposition  How can problem decomposition be efficient?

Object invention  How can learning populate world with new named objects? Object composition/decomposition?

Large-scale background knowledge  How can learners scope relevance of background concepts?

Probabilistic reasoning  How can probabilistic reasoning use single examples?
Bibliography


