Learning as Interpretation:
Human vs Statistical Learning

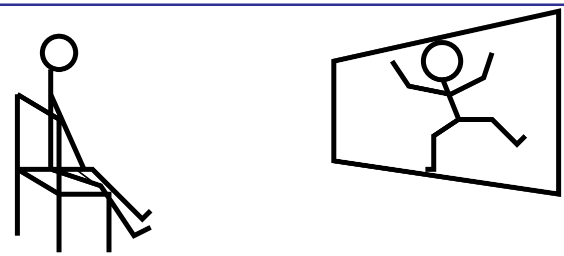
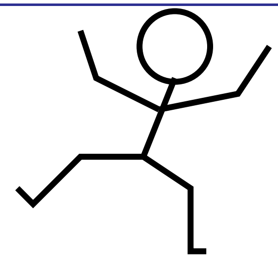
Stephen Muggleton,
Imperial College London

Human vs Statistical Learning

UK EPSRC Priority 2016-2021 - Human-like Computing

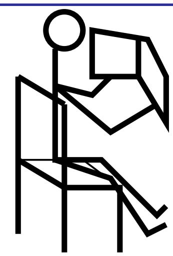
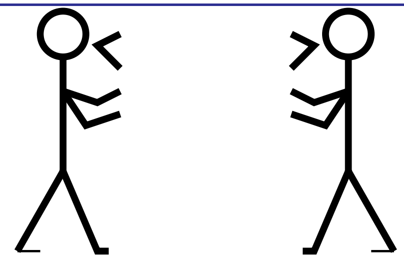
Characteristic	Human	Statistical
Examples per concept	Few (≈ 1) [Tenenbaum, 2011]	Many ($\geq 10K$)
Concepts	Many ($\geq 10K$) [Brown et al, 2008]	Few (≈ 1)
Background knowledge	Large [Brown, 2000]	Small
Structure	Modular, re-useable [Omrod et al, 2004]	Monolithic

Example 1: Dance Routine

	
Observe Visual perception	Perform Motor program

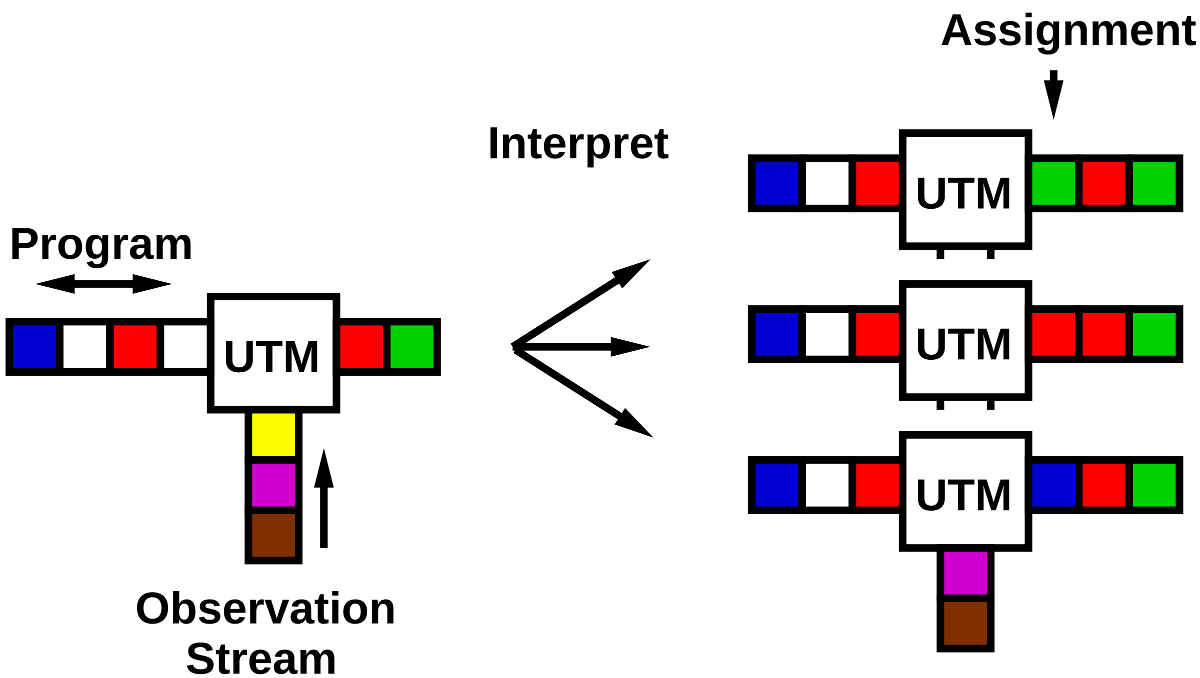
- A girl watches a dance routine on television.
- Afterwards she reproduces the routine.
- The new dance moves are incorporated into her repertoire.
- Subsequent improvisation allows re-use of parts of routines.

Example 2: Learning words in a language

	
Observe	Perform
Reading	Talking

- 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 ≈ 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

$$
\begin{array}{l}
prove([], BK, BK). \\
prove([Atom|As], BK, BK\_H) :- \\
\quad metarule(Name, MetaSub, (Atom \text{ :- } Body), Order), \\
\quad Order, \\
\quad save\_subst(metasub(Name, MetaSub), BK, BK\_C), \\
\quad prove(Body, BK\_C, BK\_Cs), \\
\quad prove(As, BK\_Cs, BK\_H).
\end{array}
$$

Metarules

Name	Meta-Rule	Order
Instance	$P(X,Y) \leftarrow$	$True$
Base	$P(x,y) \leftarrow Q(x,y)$	$P \succ Q$
Chain	$P(x,y) \leftarrow Q(x,z), R(z,y)$	$P \succ Q, P \succ R$
TailRec	$P(x,y) \leftarrow Q(x,z), P(z,y)$	$P \succ Q,$ $x \succ z \succ y$

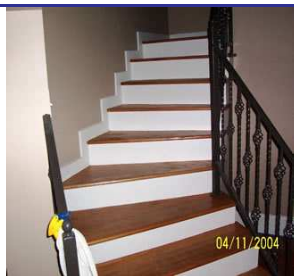
Expressivity of H_2^2

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

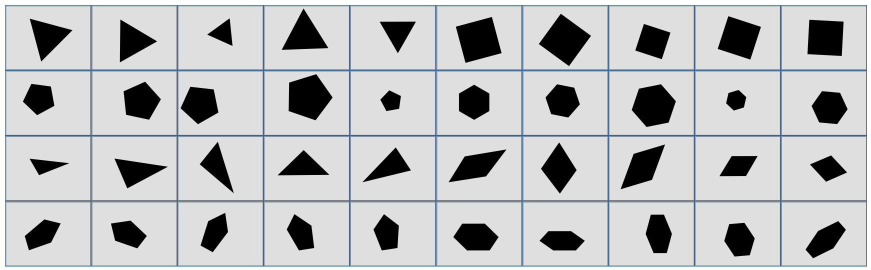
utm(S,S)	$\leftarrow$	halt(S).
utm(S,T)	$\leftarrow$	execute(S,S1), utm(S1,T).
execute(S,T)	$\leftarrow$	instruction(S,F), F(S,T).

Q: How can we limit H_2^2 to avoid the halting problem?

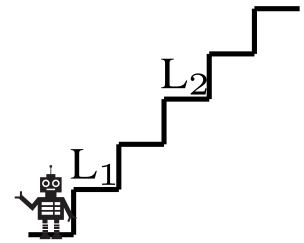
Experimental applications



Staircase



Regular Geometric



Delivery

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

- A. Cropper, S.H. Muggleton. Learning efficient logical robot strategies involving composable objects. IJCAI 2015.
- W-Z Dai, S.H. Muggleton, Z-H Zhou. Logical vision: Meta-interpretive learning for simple geometrical concepts. ILP 2015.
- S.H. Muggleton, D. Lin, A. Tamaddoni-Nezhad. Meta-interpretive learning of higher-order dyadic datalog: Predicate invention revisited. Machine Learning, 2015.
- D. Lin, E. Dechter, K. Ellis, J.B. Tenenbaum, S.H. Muggleton. Bias reformulation for one-shot function induction. ECAI 2014.