Inductive Programming
Lecture 6
Comprehensibility

Stephen Muggleton
Department of Computing
Imperial College, London and
University of Nanjing

22nd October, 2024

Papers for this lecture

Paper6.1: U. Schmid, C. Zeller, T. Besold, A. Tamaddoni-Nezhad, and S.H. Muggleton. How does predicate invention affect human comprehensibility?. Proceedings of the 26th International Conference on Inductive Logic Programming, pages 52-67, Berlin, 2017. Springer-Verlag.

Paper6.2: S.H. Muggleton, U. Schmid, C. Zeller, A. Tamaddoni-Nezhad, and T. Besold. Ultra-strong machine learning - comprehensibility of programs learned with ILP. Machine Learning, 107:1119-1140, 2018.

Motivation

- Inductive Programming
- Human feedback about induced programs
- Requires comprehensible programs
- Is program comprehensibility measureable?

Cognitive Science
Logic and Comprehensibility

- Measurements on human errors in answering questionnaires.
- Q: From the text below, is it necessarily the case that *the slithy toves did gyre*?
- **Conjunctions** easier than **Disjunctions**.

  Conj: Both *twas brillig* **and** *the slithy toves did gyre.*

  Disj: Either *twas brillig* **or** *the slithy toves did gyre.*
- **Negation** - also hard.

  NegConj: Not both *twas brillig* **and** *the slithy toves did gyre.*

Mental Model Theory

- Johnson-Laird (1983,2008) Errors - working memory overload - humans understand sentences by building semantic models.

Form	**Models**	**Load**
$p \wedge q$	$p \ q$	1
$\neg(p \vee q)$	$\neg p \ \neg q$	1
$p \oplus q$	p q	2
$\neg(p \wedge q)$	$\neg p$ $\neg q$ $\neg p \ \neg q$	3

Text comprehension tests

For many years people believed the cleverest animals after humans were chimpanzees. Now, however, there is proof that dolphins may be even cleverer than these big apes.

Question: Which animals do people think may be the cleverest?

[http://englishteststore.net]

Machine Learning and Comprehensibility

- Michie (1988) - definition of Machine Learning in terms of Predictive Accuracy and Comprehensibility.
- Mitchell (1997) - definition of Machine Learning in terms of Predictive Accuracy alone.
- Statistical Machine Learning defined in terms of Mitchell's criterion because unclear how to measure Comprehensibility.
- Use of Mechanical Turk?
- Two-way Human-Machine Learning possible?

Program comprehension tests

```
p(X,Y) :- p1(X,Z), p1(Z,Y).
p1(X,Y) :- father(X,Y).
p1(X,Y) :- mother(X,Y).
father(john,mary). mother(mary,harry).
```

Question: p(john,harry)?

Experiment 1: Effects of Predicate Invention on Comprehensibility [Paper6.1]

Predicate Invention. In the case ILP extends background knowledge B to $B \cup H$, where H is a definite program, we call predicate symbol $p \in \mathcal{P}$ an Invention iff p is defined in H but not in B.

Comprehensibility, $C(S, P)$. The comprehensibility of a definition (or program) P with respect to a human population S is the mean accuracy with which a human s from population S after brief study and without further sight can use P to classify new material sampled randomly from the definition's domain.

Experiment 1: Measureable Variables [Paper6.1, Defn 3]

Defined property	**Variable**
Comprehensibility	C
Inspection time	T
Recognition	R
Naming Time	N
Textual Complexity	Sz

Experiment 1: Experimental hypotheses

Name	**Hypothesis**
H1	$C \propto \frac{1}{T}$
H2	$C \propto R$
H3	$C \propto \frac{1}{Sz}$
H4	$R \propto \frac{1}{N}$

Experiment 1: Great-grandparent
a familiar concept

Without Invention

```
p(X,Y) :- father(X,U), father(U,Z), father(Z,Y).
p(X,Y) :- father(X,U), father(U,Z), mother(Z,Y).
p(X,Y) :- father(X,U), mother(U,Z), father(Z,Y).
p(X,Y) :- father(X,U), mother(U,Z), mother(Z,Y).
p(X,Y) :- mother(X,U), father(U,Z), mother(Z,Y).
p(X,Y) :- mother(X,U), father(U,Z), father(Z,Y).
p(X,Y) :- mother(X,U), mother(U,Z), mother(Z,Y).
p(X,Y) :- mother(X,U), mother(U,Z), father(Z,Y).
```

With Invention

```
p(X,Y) :- p1(X,U), p1(U,Z), p1(Z,Y).
p1(X,Y) :- father(X,Y).
p1(X,Y) :- mother(X,Y).
```

Experiment 1: Questionnaire - Grandparent

- What is the result of `p(bill,bob)`?
  ☐ true ☐ false ☐ don't know
- What is the result of `p(jake,harry)`?
  ☐ true ☐ false ☐ don't know
- What is the result of `p(bob,bill)`?
  ☐ true ☐ false ☐ don't know
- What is the result of `p(mary,jo)`?
  ☐ true ☐ false ☐ don't know
- What is the result of `p(john,sam)`?
  ☐ true ☐ false ☐ don't know
- What is the result of `p(X,bob)`?
  ☐ false ☐ $X =$ bill ☐ $X =$ alice
  ☐ $X =$ bill; alice ☐ don't know
- What is the result of `p(john,X)`?
  ☐ false ☐ $X =$ sam ☐ $X =$ jo
  ☐ $X =$ sam; jo ☐ don't know

Experiment 1: Results

H1	Statistically confirmed
H2	Statistically confirmed
H3	Partially confirmed
H4	Partially confirmed - recursive ancestor exception

Experiment 1: Structural identification of familiar background knowledge

```
p(X,Y)  :- p1(X,U), p1(U,Z), p1(Z,Y).
p1(X,Y) :- father(X,Y).
p1(X,Y) :- mother(X,Y).
```

parent

grandgrand parent

Experiment 2: Ultra-strong Machine Learning
Michie's Machine Learning definitions (1988)

Weak ML System uses training set to generate model with improved performance on subsequent data.

Strong ML Satisfies weak criterion and communicates model to a human in explicit form.

Ultra-Strong ML Satisfies strong criterion and model is operationally effective for humans.

Experiment 2: Additional Measureable Variables

Defined property	**Variable**
Comprehensibility of training data	C_H
Comprehensibility of ML model	C_{HM}

Experiment 2: Additional experimental hypothesis

Name	**Hypothesis**
H5	$C_H < C_{HM}$

Comprehension with and without seeing ML model

Experiment 2: Fictitious chemistry domain

Reaction observations

q1(ab,ac).	q2(aa,ac).
q1(ab,ae).	q2(aa,ae).
q1(ad,ag).	q2(ac,ag)

Test results

exothermic(ac,an).	not exothermic(aa,ab).
exothermic(aa,al).	not exothermic(ad,ai).
exothermic(ab,ag).	not exothermic(ab,aq).

Experiment 2: Fictitious Chemistry domain
an unfamiliar target

```
exothermic(X,Y) :- q1(X,Z), q1(Z,Y).
exothermic(X,Y) :- q1(X,Z), q2(Z,Y).
exothermic(X,Y) :- q2(X,Z), q2(Z,Y).
exothermic(X,Y) :- q2(X,Z), q1(Z,Y).
```

Experiment 2: Some responses

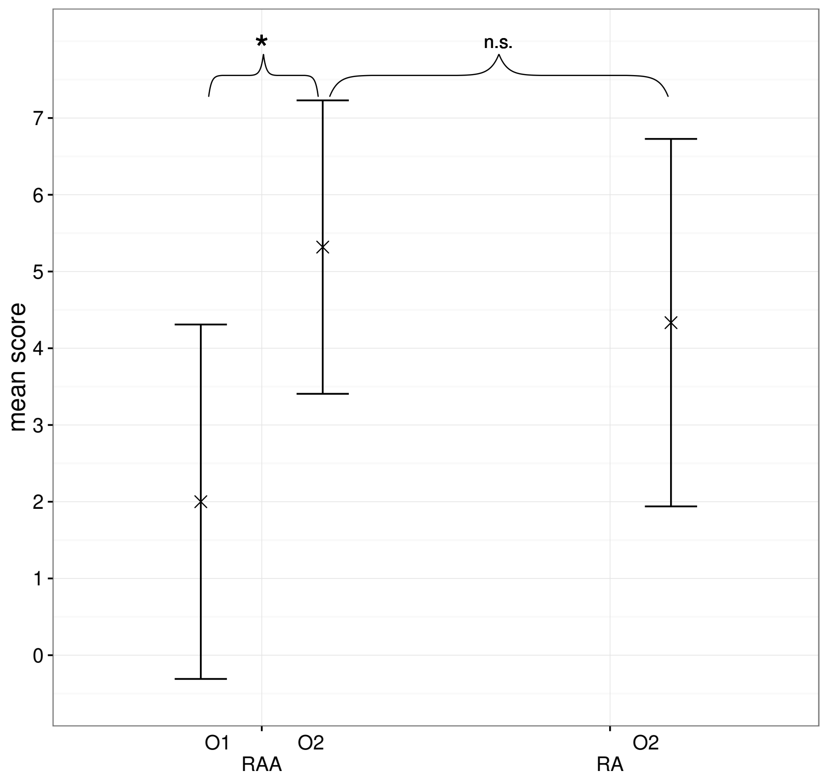
Too Specific (327)

exothermic if the substrate appears as a substrate and the product appears as a product in the same type of q. if they are both substrates or both products, or if they appear like that but in different q's, then it's not exothermic

Too General (295)

```
not_exothermic(X,Y) :- q2(X,Z), q1(Y,Z).
not_exothermic(X,Y) :- q1(X,Y).
exothermic(X,Y) :- not(not_exothermic(X,Y)).
```

Experiment 2: H5 result - Humans before and after seeing ML model [Paper6.2, Fig 8]



Experiment 2: Results [Paper6.2, Table 4]

H1	Statistically confirmed
H2	Statistically confirmed
H3	Partially confirmed
H4	Partially confirmed
H5	Statistically confirmed

Summary

- Human feedback about induced programs
- How do we measure comprehensibility?
- Johnson-Laird's Mental Model Theory
- Comprehension tests for text
- Comprehension tests for logic programs
- Testing properties of comprehensible theories
- Experiment 1 - familiar concepts - kinship
- Testing Michie's Ultra-Strong Machine Learning
- Experiment 2 - unfamiliar concepts - exothermic
- Result - Machines can teach Humans unfamiliar concepts