Inductive Programming: Tutorial 2
Domain-Specific Languages and Background Knowledge

Stephen Muggleton

The aim of this tutorial is to help you understand concepts in Lecture 2, involving Domain-Specific Languages and Background Knowledge

Question 1

Describe Valiant's PAC-learning model.

Solution Valiant's PAC-learning model defines a class of polynomial time learning algorithms which, when given sufficient training examples, have high Probability of choosing a hypothesis which is Approximately Correct on unseen examples.

Question 2

Assume a PAC algorithm A is applied to two hypothesis spaces of sizes a) 10^5 and b) 10^2. According to the Blumer bound, what is the minimum number of randomly sampled examples A will require to guarantee with probability 0.7 that A's error will be at most 30%.

Solution $\delta = \epsilon = 0.3$. The Blumer bound is $m \geq \frac{(ln|\mathcal{H}|+ln\frac{1}{\delta})}{\epsilon}$. The two cases described are as follows.

a) Subsituting into the Blumer Bound gives.

$$\begin{aligned} m &\geq \frac{(ln|\mathcal{H}| + ln\frac{1}{\delta})}{\epsilon} \\ &\geq \frac{(ln(10^5) + ln\frac{1}{0.3})}{0.3} \\ &\geq \frac{(11.51 + 1.20)}{0.3} \\ &\geq 43 \end{aligned}$$

b) Subsituting into the Blumer Bound gives.

$$\begin{aligned} m &\geq \frac{(4.61 + 1.20)}{0.3} \\ &\geq 20 \end{aligned}$$

Question 3

What are the relationships between Background Knowledge, Metarules and Examples in MIL?

Solution Given input (B, M, E^+, E^-) where background B is a logic program, metarules M are higher-order clauses and examples E^+, E^- are ground atoms. An MIL algorithm returns a logic program hypothesis H such that $M \models H$ and $H \cup B \models E^+$ and $H \cup B \not\models E^-$.

Question 4

What is

a) the H^2_2 hypothesis space?

b) the size of the H^2_2 hypothesis space?

c) the order of the sample complexity based on the Blumer bound?

Solution

a) The H^2_2 hypothesis space consists of definite clauses with at most two body atoms and at most predicate arity of two.

b) H^2_2 hypothesis space is $\mathcal{H}$ is $O(|M|^n p^{3n})$ given M metarules, n clauses in the hypothesis and p predicate symbols.

c) For fixed M, p we have m is $O(\frac{n}{\epsilon})$.

Question 5

Describe three techniques implemented in Metagol to increase efficiency of the search.

Solution

1) Ordering the Herbrand Base guarantees termination of derivations. Orderings can be either lexicographic or interval based.

2) Episodes allow sequence of related learned concepts, which reduces the complexity of the search from $\prod_i |H_i|$ to $\sum_i |H_i|$.

3) Iterative deepening search $H_0, .., H_n$ returns $h_n \in H_n$ where n minimal. Hypotheses in H_i have i clauses. Search minimal. Returned consistent hypothesis minimal number of clauses.