CSE 513T Theory of Artificial Intelligence & Machine Learning November 8th, 2016

Lecture 20: Well-Founded Semantics

Lecturer: Brendan Juba *Scribe: Jacob Solawetz*

The last day in which we thought the world made sense.

Today we covered the following topics:

1) Partial Models
2) Quotient Operators
3) Well-Founded Semantics

Last time we learned about Negation as Failure 'naf'. 'naf' is a way to reason using lack of contrary evidence, and can be a way to think about dealing with common sense. Here is a friendly example using naf reasoning that we might use in every day thinking:

$$[bird \wedge not(\neg fly)] \rightarrow fly$$

We can think of this as 'default' reasoning. There are KBs that can be rather non-intuitive using NAF reasoning. We saw last time the following KB in particular:

$$[notx_1] \rightarrow x_2$$
$$[notx_2] \rightarrow x_1$$
$$[notx_2] \rightarrow x_1$$
$$[x_1] \rightarrow x_3$$

Recall that "not" means the following literal is considered unprovable.

The way out of these kind of messes is to withhold judgement on whether or not facts are provable. In any kind naf reasoning we would like to see the following conditions:

1) Anything that can be proved without naf should still be provable.
2) We want consistent judgements
3) Don't conclude circular justifications
4) Want to recongnize (as often as possible) when something is not possible. That is, our naf should do some reasoning.

Definition: Partial Model

A partial model is given by disjoint subsets of literals, L_T (provable) and $L_\perp$ (not provable) such that for every rule in the KB

$$[l_1 \wedge ...l_r \wedge notl_{r+1}, ..., notl_s] \rightarrow l^*$$

if $[l_1 \wedge ...l_r \wedge notl_{r+1}, ..., notl_s] \in L_T$ then $l^* \in L_T$ and if $l^* \in L_\perp$ then at least one of the literals must be $[l_1 \vee ...l_r \vee notl_{r+1}, ..., notl_s] \in L_\perp$.

Note: we treat facts as rules with empty bodies, so these all must be in L_T.

First example of a partial model. Treat $notl_i$ as new literals, put everything not provable by chaining into $L_\perp$. This satifies the above definition but it does poorly on condition (4), above. We want our partial model to do some reasoning.

Consider a better model

Definiton: A least partial model w.r.t. a set M of partial models is a partial model $(L_T^*, L_\perp^*) \in M$, s.t for every other partial model $(L_T', L_\perp') \in M$, $L_\perp' \subseteq L_\perp^*$ **and** $L_T^* \subseteq L_T'$

Note $L_\perp' \subseteq L_\perp^*$ helps ensure condition 4. and $L_T^* \subseteq L_T'$ helps insure condition 3.

Theorem: The partial model obtained by taking L_T^* to be the set of all literals provable by chaining from the KB when we treat 'not l' as new literals, and taking $L_\perp^*$ to be the set of literals s.t. if all occurences of 'not l' are dropped from rules of the KB, remain unprovable, is the least partial model among those for which we do not place any 'not l' in $L_\perp$. (Note we generally disallow naf'd literals in the heads of rules).

Question: What is the least partial model for the following KB:

$$[notx_1] \rightarrow x_2$$
$$[notx_2] \rightarrow x_1$$
$$[notx_3] \rightarrow x_3$$

Answer:

$$L_T^* = \emptyset$$
$$L_\perp^* = \{\neg x_1, \neg x_2, \neg x_3\}$$

The answer is obtained by applying the preceding theorem.

Proof of Theorem: We first note that whenever a least partial model exists it must be unique. Indeed if $(L_T', L_\perp')$ and $(L_T'', L_\perp'')$ were both least partial models then $L_\perp' \subseteq L_\perp' \subseteq L_\perp''$ so $L_\perp' = L_\perp''$. We reason similarly for L_T', L_T''. Uniqueness was easy. To prove the rest we set out a couple of claims.

Claim 1: Every L_T for such a partial model must contain L_T^*

Proof of Claim 1: By induction on the length of the shortest chaining derivation establising each literal in L_T^*. If literal l is in L_T^* due to a t + 1 long chaining proof, then each literal l' in the body of the rule must have a length $\leq t$ chaing derivation. By the IH, since each such literal $l' \in L_T^*$. Now, since every literal l' in the body of this rule is in L_T', by the definition of a partial model $l \in L_T'$. □

Claim 2: "My unprovable set is bigger than yours", $L_\perp' \subseteq L_\perp^*$

Proof of Claim: By contradiction. If $L_\perp' \nsubseteq L_\perp^*$, there is some $l \in L_\perp'$ such that when the occurences of 'not l' are dropped from the KB, there is a chaing proof of l. Consider the literal l^* with the shortest chaing proof. By definition of parital model, since $l^* \in L_\perp'$ for every rule with l^* in the head there must have been some other literal l' in the body that is also in $L_\perp'$, that cannot have been one of the naf'd literals (by assumption). Since i the chaining proof of l^* there must have been a proof of such 'l' that is shorter than the proof of l^*, a contradiction with our choice of l^*.

Now we have only left to verify that the theorem produces a partial model. To see this, note that in the modified KB in which naf'd literals are dropped, each literal either has a chaining proof or it doesn't. If l appears in the head of any rule in which every literal in the body is provable l is also regarded as provable. Contrapostively, any $l \in L_\perp^*$ mus tin every rule of this modified KB, have some literal in the body that is also not provable. So $L_\perp^*$ in particular is needed for a partial model. □

Definition: Given a pair of sets of literals $(L_T, L_\perp)$ and a KB, the **quotient of KB modulo $(L_T, L_\perp)$** is obtained as follows:

1) Substitute 1 for occurences of 'not l' for $l \in L_\perp$
2) Subsitute 0 for occurrences of 'not l' for $l \in L_T$
3) Simplify resulting KB by deleting rules in which 0 appears in the body and omitting occurrences of 1 in rule bodies (aka clean up the KB)

Example Suppose our KB is:

$$x_1$$
$$[notx_1] \rightarrow x_2$$
$$[notx_3] \rightarrow x_3$$

Then $L_T^* = \{x_1\}$ and $L_\perp^* = \{\neg x_1, \neg x_2, \neg x_3\}$. Then taking the quotient yields the new KB:

$$x_1$$
$$[notx_3] \rightarrow x_3$$

We could do this process repeatedly and it appears that repeated quotients will eventually converge. Indeed this is a main theorem of Well founded semantics.

Theorem: Consider the sequence $(L_T^0, L_\perp^0) = (\emptyset, \emptyset)$ and $(L_T^i, L_\perp^i)$ are the partial model of the quotient of the KB with $L_T^{i-1}, L_\perp^{i-1})$ (that does not pace any 'not l' in the $l_\perp$). Then this sequence converges to its well founded Semantics $(L_T^*, L_\perp^*)$ s.t. for each of the standard literals (non naf'd) the following are true:

1) if there is a chaining proof of l in KB/$(L_T^*, L_\perp^*)$, then $l \in L_T^*$
2) If $l \in L_\perp^*$ then there is no chaing proof of l in KB/$(L_T^*, L_\perp^*)$

Furthermore, there is an algorithm to compute $(L_T^*, L_\perp^*)$ in time $O(n^2 \mid KB \mid)$

Proof next class.