Logics for Artificial Intelligence- AI Master

Session 4 – Model Theory in First-Order Logic

The validation method seen in the previous weeks in First-Order Logic (Resolution) belongs to **Proof Theory**, as we try to use some syntactic inference rules in order to build a formal proof. In this session we study validation methods from **Model Theory,** in which we examine all the possible situations in which the world may be so that we can discover if an argument is *valid* (by checking if, in all the situations in which the premises hold, the conclusion is also true) or *invalid* (if there is at least one possible case in which the premises are true but the conclusion is false).

A possible situation will be represented with an **interpretation**. In First-Order Logic an interpretation has the following components:

- **Domain**: non-empty set of elements.
- **Interpretation of the constants**: function that assigns, to each constant, an element of the domain.
- **Interpretation of the functions**: for each *n*-ary function, the interpretation has to include a total function that, given *n* elements of the domain, returns an element of the domain.
- **Interpretation of the predicates**: for each *n*-ary predicate, the interpretation has to include a total function that, given *n* elements of the domain, returns true/false depending on whether those *n* elements are related by the predicate.

Example: imagine a situation in which we have formulas containing the constants *a* and *b*, a binary function *f* and a unary predicate *P*. A possible interpretation could be the following:

- Domain = {i1, i2, i3}
- Interpretation of the constants: a => i2, b => i3.
- Interpretation of the functions: f(i1,i1)=i2; f(i1,i2)=i1; f(i1,i3)=i2; f(i2,i1)=i1; f(i2,i2)=i2; f(i2,i3)=i1; f(i3,i1)= i1; f(i3,i2)=i1; f(i3,i3)=i2.
- Interpretation of the predicates: P(i1)=true; P(i2)=false; P(i3)=true.

We can check if a formula is valid in an interpretation by following these steps:

- Replace the constants by their associated domain elements.
- Replace the universal quantifiers by a conjunction on all the elements of the domain, and the existential quantifiers by a disjunction on all the elements of the domain.
- Apply the functions to obtain their results (domain elements), given the interpretation of the functions.
- For every predicate applied to domain elements, check if the relationship holds, using the interpretation of the predicates.
- Use the truth tables of the logical connectives to find the final truth value of the whole formula.

The truth tables of the basic logical connectives are the following:

P	Q	¬P	(P^Q)	(PvQ)	(P→Q)
False	False	True	False	False	True
False	True	True	False	True	True
True	False	False	False	True	False
True	True	False	True	True	True

Example: we can compute the truth value of the formula ∃x P(f(x,a)) in the previous interpretation.

- The constant *a* is replaced by its interpretation (i2) (∃x P(f(x,i2))).
- The existential quantifier is replaced by a disjunction on all the elements of the domain (P(f(i1,i2)) ˅ P(f(i2,i2)) ˅ P(f(i3,i2))).
- We apply the interpretation of the function f (P(i1) ˅ P(i2) ˅ P(i1)).
- We check the interpretation of the predicate P (true ˅ false ˅ true).
- Applying the truth table of disjunction we obtain the final value (true).

A formula is valid (**tautology**) iff it is true in all interpretations. A formula is unsatisfiable (**contradiction**) iff it is false in all interpretations. A formula that is neither valid nor unsatisfiable is called **contingent**. An interpretation is a **model** of a formula iff the formula is true in that interpretation, and a **counter-example** iff the formula is false.

Using model theory, we can check whether B is a logical consequence of a set of premises A = {A_1 ^ … ^ A_n} in any of the two following ways:

- Checking if A→B is valid.
- Checking if A^ ¬B is unsatisfiable.

We can check if a formula is valid with the following procedure:

1. Consider a domain of a single element, with size s=1.
2. Consider all the possible interpretations for domains with size s.
3. Check if the formula is valid on all the interpretations. If it is false in some interpretation, we have found a counter-example that shows that the formula is not valid, and we can stop; otherwise, we proceed to the next step.
4. Consider a domain of size s=s+1. Go back to step 2.

This procedure can show in a finite time that a formula is invalid, but it can't show that a formula is valid. In order to do that, we would have to do some kind of inductive proof that showed that, if the formula is true for all interpretations for domains of size *n*, it will also be true in all interpretations of size *n+1*.

We could also define in an analogous way how to check the unsatisfiability of a formula.

Proof Theory and Model Theory are linked by the **soundness** (all theorems are tautologies) and **completeness** (all tautologies are theorems) of Predicate Logic[1]. If we want to prove that an argument is valid, we would normally use proof theory; if we want to show that an argument is invalid, we would normally use model theory to look for a counter-example.

Complementary material you should study on week 5:

- Textbook by Brachman and Levesque: sections 2.3 and 2.4.
- Stanford course by Genesereth and Rao: sections 2.3, 2.4 and 2.5 (propositions), 6.3, 6.4 and 6.5 (predicates) and 7.2.

[1] A theorem is a formula that may be deduced from the empty set of premises, using Natural Deduction or Resolution.