Lecture notes 20.0

Herbrand models, substitutions

COMP 2411, session 1, 2004

Names for individuals

We know that an interpretation over a vocabulary has a *domain*, or a set of *individuals*. Some of these individuals have no *name*, *i.e.*, interpret no closed term.

For instance, consider a vocabulary with a unary predicate symbol p as only nonlogical symbol. There are many models of $\forall x\, p(x)$. There are models of $\forall x\, p(x)$ of cardinality n for all $n > 0$. There are infinite models of $\forall x\, p(x)$.

All the models of $\forall x\, p(x)$ have the particularity that their members have no name: there are no closed terms over the vocabulary.

Nice interpretations have the property that each of their individuals interprets a unique closed term.

Example

Suppose that the vocabulary contains two constants tom and ann and no other function symbol. A nice interpretation would just have two elements; one would interpret tom, the other would interpret ann.

Suppose that the only function symbols of the vocabulary are a constant a and a unary function symbol s. A nice interpretation would have infinitely many elements. Every element would be interpret a unique term of the form $\overbrace{s(\ldots s(}^{n} a) \ldots)$, abbreviated as $s^n(a)$. For example, we can choose a structure $\mathfrak{M}$ with $|\mathfrak{M}| = \mathbb{N}$ and decide that for all $n \in \mathbb{N}$, n interprets $s^n(a)$.

We will define a class of nice interpretations called Herbrand interpretations.

Nice interpretations

If the vocabulary contains n constants and no function symbol of nonnull arity, then all nice interpretations have n elements.

If the vocabulary contains a constant and a function symbol whose arity is nonnull, then all nice interpretations are infinite.

If the vocabulary contains no constant, then there is no nice interpretation over the vocabulary, even if the vocabulary contains function symbols of nonnull arity.

Suppose that the vocabulary consists of a constant a, a unary function symbol s, and a unary predicate p. Then $P = \{\exists x\, p(x), \neg p(a), \forall x(\neg p(x) \rightarrow \neg p(s(x))\}$. Then P is satisfiable, but has no nice interpretation.

Herbrand universe and base

Consider a vocabulary containing at least one constant.

The set U of all closed terms over the vocabulary is called the Herbrand universe.

The set B of all closed atomic formulas is called the Herbrand base.

We also talk about Herbrand universe and base for a set of formulas P containing at least one constant: in that case, the (nonlogical part) of the vocabulary is defined as the set of nonlogical symbols occurring in P. We will use the notation U_P and B_P, respectively.

Examples

Consider the set of formulas P:

$\mathrm{odd(s(0))}.$
$\mathrm{\forall x(odd(x) \rightarrow odd(s(s(x))))}.$

Then $U_P = \{0, s(0), s(s(0)), s(s(s(0))), \ldots\}$ and
$B_P = \{\mathrm{odd(0)}, \mathrm{odd(s(0))}, \mathrm{odd(s(s(0)))}, \mathrm{odd(s(s(s(0))))}, \ldots\}$.

Consider the set of formulas P:

$\mathrm{owns(owner(corvette), corvette)}.$
$\mathrm{\forall x(owns(x, corvette) \rightarrow happy(x))}.$

Then
$U_P = \{\mathrm{corvette}, \mathrm{owner(corvette)}, \mathrm{owner(owner(corvette))}, \ldots\}$
and $B_P = \{\mathrm{owns(s,t)} \,|\, \mathrm{s,t} \in \mathrm{U_P}\} \cup \{\mathrm{happy(s)} \,|\, \mathrm{s} \in \mathrm{U_P}\}$.

Herbrand interpretations (1)

Herbrand interpretations could be defined as interpretations $\mathfrak{M}$ all of whose individuals have a unique name. But if $|\mathfrak{M}|$ ($\mathfrak{M}$'s domain) is in one-to-one correspondence with the set of closed terms, we can just assume that $|\mathfrak{M}|$ is the set of closed terms. Moreover, we assume that all closed terms are interpreted canonically. Formally:

Given a set of formulas P, a *Herbrand interpretation of* P is an interpretation $\mathfrak{M}$ over the language of P such that:

- $|\mathfrak{M}| = U_P$.
- For every n-ary function symbol f occurring in P, $f^{\mathfrak{M}}$ is such that for all $a_1, \ldots, a_n \in |\mathfrak{M}|$, $f^{\mathfrak{M}}(a_1, \ldots, a_n) = f(a_1, \ldots, a_n)$.

In particular, for all constants c occurring in P, $c^{\mathfrak{M}} = c$.

Herbrand interpretations (2)

Note that if p is an n-ary predicate symbol occurring in P, then $p^{\mathfrak{M}}$ is a subset of U_P^n.

Suppose that the only predicate symbol in P is $p/1$. Then it is natural to identify a Herbrand interpretation with a subset of the Herbrand universe.

More generally, a Herbrand interpretation can be identified with a set of labelled n-tuples over the Herbrand universe, where the label is a predicate symbol of arity n.

It is then natural to consider the intersection of two Herbrand interpretations, or their union, etc.

Example

Consider again the set of formulas P:

$\mathrm{odd(s(0))}$.
$\forall \mathrm{x}(\mathrm{odd(x)} \rightarrow \mathrm{odd(s(s(x)))})$.

Possible Herbrand interpretations are:

- $\mathfrak{M}_1 = \emptyset$.
- $\mathfrak{M}_2 = \{\mathrm{odd(s(0))}\}$.
- $\mathfrak{M}_3 = \{\mathrm{odd(s(0))}, \mathrm{odd(s(s(0)))}\}$.
- $\mathfrak{M}_4 = \{\mathrm{odd(s^n(0))} \mid \mathrm{n} \in \{1, 3, 5, \ldots\}\}$
- $\mathfrak{M}_5 = B_P$.

Obviously, only $\mathfrak{M}_4$ and $\mathfrak{M}_5$ are models of P.

Herbrand models

Given a set P of formulas, a *Herbrand model of* P is by definition a Herbrand interpretation that is a model of P.

We have seen that some sets of formulas have a model (are satisfiable), but have no Herbrand model.

Thanks to Skolemization, we know how to transform a formula φ into a set X of clauses such that φ is satisfiable iff X is satisfiable. With that in mind, we give the following result.

Proposition: Every satisfiable set of clauses has a Herbrand model.

Less formally, every set of clauses has a 'nice' model, all of whose individuals have a unique name.

Proof (1)

Let set of clauses P be satisfiable. Choose a model $\mathfrak{M}$ of P.

Let $\mathfrak{N}$ be the Herbrand interpretation identified with $\{F \in B_P \mid \mathfrak{M} \models F\}$. We show that $\mathfrak{N} \models P$.

Suppose for a contradiction that $\mathfrak{N}$ is not a model of P. Choose a clause $C \in P$ such that $\mathfrak{N} \not\models C$.

C is of the form $L_0 \vee \ldots \vee L_n$ where for all $i \leq n$, L_i is atomic or the negation of an atomic formula. Assume for instance that $C = p(x, y) \vee \neg q(x) \vee \neg p(y, z) \vee r(x, z)$.

Then $\mathfrak{N} \not\models C$ means that
$\mathfrak{N} \models \exists x \exists y \exists z (\neg p(x, y) \wedge q(x) \wedge p(y, z) \wedge \neg r(x, z))$.

Hence we can find $a, b, c \in |\mathfrak{N}|$ such that
$\mathfrak{N} \models \neg p(\bar{a}, \bar{b}) \wedge q(\bar{a}) \wedge p(\bar{b}, \bar{c}) \wedge \neg r(\bar{a}, \bar{c})$.

Proof (2)

But $\mathfrak{N}$ is a Herbrand interpretation, hence a_1, a_2, a_3 have names—respectively some closed terms t_1, t_2, t_3—and $\mathfrak{N} \models \neg p(t_1, t_2) \wedge q(t_1) \wedge p(t_2, t_3) \wedge \neg r(t_1, t_3)$.

Note that $p(t_1, t_2)$, $q(t_1)$, $p(t_2, t_3)$ and $r(t_1, t_3)$ are all members of B_P.

By the definition of $\mathfrak{N}$, it follows that
$\mathfrak{M} \models \neg p(t_1, t_2) \wedge q(t_1) \wedge p(t_2, t_3) \wedge \neg r(t_1, t_3)$.

Hence $\mathfrak{M} \not\models \forall x \forall y \forall z (p(x, y) \vee \neg q(x) \vee \neg p(y, z) \vee r(x, z))$.

Hence $\mathfrak{M} \not\models C$. Contradiction.

The reasoning we did using this particular clause can be applied mutatis mutandis to an arbitrary clause, which completes the proof of the proposition.

Corollary

Corollary: Let φ be a formula and let X be a set of clauses obtained from $\neg\varphi$ by Skolemization. Then φ is valid iff X has no Herbrand model.

This is good news for two reasons:

- We have fewer structures to deal with.
- Herbrand interpretations look nicer than interpretations that are not Herbrand. So the task of proving that φ is valid should be easier working from X than working from φ: we indeed only have to derive a contradiction from closed instances of members of X.

Still generating all closed instances of a set of clauses would be inefficient. In order to get a more efficient proof procedure, we need to introduce the notion of substitution and then of unifier.

Substitutions

In this part we use upper-case letters to represent variables and easily distinguish them from constants.

A substitution (for a vocabulary $\mathcal{V}$) is a mapping from a finite set of variables to the set of terms over $\mathcal{V}$.
Let distinct variables $X_0, \ldots, X_n$ and terms $t_0, \ldots, t_n$ be such that $t_i \neq X_i$ for all $i \leq n$. The substitution that maps X_i to t_i for all $i \leq n$ is represented as $\{X_0/t_0, \ldots, X_n/t_n\}$.

The conditions that $X_0, \ldots, X_n$ are pairwise distinct and that $t_i \neq X_i$ for all $i \leq n$ guarantee that is no inconsistency, redundancy, or empty information in the notation $\{X_0/t_0, \ldots, X_n/t_n\}$.

Domain and range

The empty substitution is denoted ϵ.

The *domain* of a substitution θ, *i.e.*, the set of variables on which θ is defined, is denoted $Dom(\theta)$. Hence $Dom(\{X_0/t_0, \ldots, X_n/t_n\}) = \{X_0, \ldots, X_n\}$.

The *range* of a substitution $\theta = \{X_0/t_0, \ldots, X_n/t_n\}$ is defined here as the set of all variables that occur in some of the terms $t_0, \ldots, t_n$. It is denoted $Range(\theta)$.

For instance, $Range(\{X/f(Z), Y/g(a, X)\}) = \{Z, X\}$.

(Note that the previous definition is not standard; for the standard definition, the range of $\theta = \{X_0/t_0, \ldots, X_n/t_n\}$ is $\{t_0, \ldots, t_n\}$.)

Application

Consider a substitution $\theta = \{X_0/t_0, \ldots, X_n/t_n\}$. Let E be a term or a formula without quantifiers. The *application of* θ *to* E, denoted $E\theta$, is the sequence of symbols obtained from E by replacing *simultaneously* every occurrence of X_i by t_i, for all $i \leq n$.

Note that $E\theta$ is a term if E is a term, and that $E\theta$ is a formula if E is a formula.

Note that in particular, $X\theta = t$ if $X/t \in \theta$, and $X\theta = X$ otherwise.

We call $E\theta$ an *instance of* E (w.r.t. θ).

Note that every term is an instance of itself w.r.t. ϵ.

Examples

Note that for $\theta = \{X_0/t_0, \ldots, X_n/t_n\}$, $E\theta$ is just another notation for what we have denoted by $E[t_0/X_0, \ldots, t_n/X_n]$.

- $p(f(X), g(A,Y), h(A,X,Z))\{X/f(f(Y)), U/f(a)\} = p(f(f(f(Y))), g(A,Y), h(A, f(f(Y)), Z))$
- $p(X) \vee r(X,Z) \vee q(X,Y,Z)\{X/a, Y/b\} = p(a) \vee r(a,Z) \vee q(a,b,Z)$
- $p(f(X,Z), f(Y,a))\{X/a, Y/Z, W/b\} = p(f(a,Z), f(Z,a))$
- $p(X,Y)\{X/f(Y), Y/b\} = p(f(Y), b)$
- $p(X,Y,Z)\{X/f(Y), Y/a\} = p(f(Y), a, Z)$

Composition (1)

Given two substitutions θ and σ, the composition of θ and σ, *i.e.*, the function $\sigma \circ \theta$, is denoted $\theta\sigma$.

The notation makes sense since the argument of $\sigma \circ \theta$ is on the right hand side, whereas the argument of $\theta\sigma$ is on the left hand side: for all $X \in Domain(\sigma \circ \theta)$, $(\sigma \circ \theta)(X) = X(\theta\sigma)$.

For example:

- $\{X/f(Z), Y/g(a,X)\}\{Z/g(X,b), U/f(c)\} = \{X/f(g(X,b)), Y/g(a,X), Z/g(X,b), U/f(c)\}$
- $\{X/f(Z), Y/W\}\{X/a, Z/a, W/Y\} = \{X/f(a), Z/a, W/Y\}$
- $\{X/f(X), Y/g(a,T), Z/b\}\{X/a, Y/b, V/X\} = \{X/f(a), Y/g(a,T), Z/b, V/X\}$

Composition (2)

Let $\theta = \{X_0/s_0, \ldots, X_m/s_m\}$ and $\sigma = \{Y_0/t_0, \ldots, Y_n/t_n\}$ be two substitutions.

$X_0, \ldots, X_m$ are pairwise distinct, $Y_0, \ldots, Y_n$ are pairwise distinct, but some X_i might be equal to some Y_j for some $i \leq m$ and $j \leq n$.

It can be verified that $\theta\sigma$ can be obtained as follows.

- First, write down $\{X_0/s_0\sigma, \ldots, X_m/s_m\sigma, Y_0/t_0, \ldots, Y_n/t_n\}$
- Then remove all Y_j/t_j such that $Y_j \in \{X_0, \ldots, X_m\}$, for $j \leq n$.
- Then remove all $X_i/s_i\sigma$ such that $X_i = s_i\sigma$, for $i \leq m$.

Properties of substitutions (1)

A substitution θ is said to be *idempotent* iff $\theta = \theta\theta$. So given a term or a formula without quantifiers, say E, applying θ to E once or many times has the same effect on E.

It is immediately verified that a substitution θ is idempotent iff $Dom(\theta)$ and $Range(\theta)$ are disjoint.

Also, substitutions satisfy all the properties of mappings in general:

- Associativity: $\theta(\sigma\tau) = (\theta\sigma)\tau$
- The empty substitution ϵ is a neutral element: $\epsilon\theta = \theta\epsilon = \theta$.

Properties of substitutions (2)

But in general substitutions do not satisfy the commutativity property: $\sigma\theta$ can be different from $\theta\sigma$, as shown by the following example.

$\{X/f(Y)\}\{Y/a\} = \{X/f(a), Y/a\} \neq$
$\{Y/a\}\{X/f(Y)\} = \{Y/a, X/f(Y)\}.$

Also, substitutions usually do not have an inverse: given a substitution θ, there is usually no substitution σ such that $\theta\sigma = \epsilon$.

In particular, whenever θ contains an element of the form X/t where t is not a variable, θ does not have an inverse.