AUTOMATED REASONING

SLIDES 4:

FORMAL NOTIONS

Structures and Models
Herbrand Interpretations and Models
Soundness and Completeness Properties
Soundness of Resolution
Completeness of Resolution (Outline)

KB - AR - 12

Properties of Inference Systems: 4ai

Slides 4 include some material on the properties of inference systems, including material on first order structures. The notion of a Herbrand interpretation, a first order structure with a very particular domain, is introduced and it's explained why Herbrand Interpretations are important for soundness and completeness of resolution. The "Useful Theorem" on Slide 4bii and the Skolemisation property on 4di capture this. These properties mean that when proving theorems about resolution it is sufficient to restrict considerations to Herbrand interpretations only, substantially simplifying the proofs. Also, when using refutation as a proof technique to show (un)satisfiability of data, it is sound to consider the clausal form representation of the data.

The proofs of the theorems in Slides 4 (if not given here) can be found either in Appendix1 or in Chapter Notes 1 at www.doc.ic.ac.uk/~kb. ***Note that none of Appendix1 is examinable***.

Don't worry if you're not familiar with first order structures, as it is sufficient to understand about Herbrand structures. However, if you are familiar, you can check your understanding using the clauses on Slide 4aiii: $\forall x[(P(x) \rightarrow P(f(x))]$, P(a) , P(b), Q(c, g(c)) with the structure on the slide.

Take Domain = Lists over the (English) alphabet and the following mapping of terms to Domain
a is "a", b is "the", c is "hit"
f(x) is the word formed by appending 's" to x
P(x) is true if x is a correct English word

a) Which of $\forall x\, (P(x) \rightarrow P(f(x))$, P(a) or P(b) are true in this structure?
b) Choose interpretations for g and Q that make Q(c,g(c)) true in the structure.

The notation $val(S)_I[x/d]$ may be used to denote the valuation of the sentence S in the structure I in which free occurrences of x in S are replaced by Domain element d. For example, $val(P(x))_I[x/'a']$ means " val ((interpretation of P in I)('a'))", which for the structure I given above is " val('a' is a correct English word)". Note that $\forall x.S$ is true in structure I if $val(S)_I[x/d]$ is true in I for <u>every</u> d in the Domain, and $\exists x.S$ is true in structure I if $val(S)_I[x/d]$ is true in I for <u>some</u> d in the Domain.

Structures 4aii

• (First order) sentences are written in a language L, which uses predicates and terms constructed from names in the signature Sig(L) = <P, F, C>, where P = predicates, F = function symbols, C = constants.

Example: Let Sig(L) = <{P,Q},{f},{a,b}>
S = {$\forall x\, (P(x) \rightarrow P(f(x)))$, P(a), P(b), Q(a, f(a))}
a and b are constants, f has arity 1, P has arity 1 and Q has arity 2

• A structure I for L (sometimes informally called an interpretation) consists of
a non-empty domain D, and
an interpretation (i.e. a meaning) for each symbol in Sig(L):

• $c \in C$ is interpreted by an element I(c) of D
• f (of arity n) $\in$ F is interpreted by a function I(f) of arity n from D^n to D
• p (of arity m) $\in$ P is interpreted by a relation I(P) of arity m on D^m

Structure 1: Domain = {integers} = Int

- I(a) = 0, I(b) =2
- I(f) is the function Int -> Int, where x -> x+2 (i.e. the "add 2" function)
- I(P) = even (ie P(x) is true iff x is even)
- I(Q) = less (ie Q(x,y) is true iff x is less than y)

Structures (continued) 4aiii

Structure 1 again: Domain = {integers}
I(a) = 0, I(b) =2 I(f) is the function x -> x+2
I(P) = even I(Q) = less

The interpretation I is extended to apply to all ground terms in language L:

I(f(t1, ..., tn)) = I(f)(I(t1), ..., I(tn)) for a functional term of arity n

Also, I(x) =x for a bound variable x

I(P) allows to give a valuation for all ground atoms

val(P(t1, ..., tn)) = val(I(P)(I(t1), ..., I(tm))) for an atom of arity m

val(P(a)) = val(even(0)) = True
val(P(b)) = val(even(2)) = True
val($\forall x\, (P(x) \rightarrow P(f(x))$) = val($\forall x\, (\text{even}(x) \rightarrow \text{even}(x+2))$) = True

Structure 2: a Herbrand Structure

4aiv

Given: Sig(L) = <{P},{f},{a,b}> S = {∀x (P(x) → P(f(x))), P(a) , P(b)}

A Herbrand Structure for Sig(L):

- Domain ={*a,b,c, f(a), …, f(f(a)),…,f(f(f(a))),…*}
  (i.e. the set of names of terms in L, sometimes written as a,b, etc.)
- I(a) = *a*, I(b) = *b*, I(f) = *f*, hence I(f(a)) = I(f)(I(a)) = *f(a)*
  i.e. domain elements are, in effect, mapped to (interpreted as) themselves

NOTE: mapping of constants and functors is fixed in a Herbrand structure

- I(P) = *P*

The valuation of atoms can be given explicitly:

e.g. val(P(a)) = val(I(P)(I(a))) = val(*P(a)*) = True (say)

Assign True to atoms of the form $P(f^n(a))$ and False to atoms of the form $P(f^n(b))$

i.e. *P(a) = P(f(a)) = P(f(f(a)))* =… = True and *P(b) = P(f(b))* = … = False

A ***Herbrand interpretation*** is represented by a subset of the set of atoms:
e.g. { P(a), P(f(a)), P(f(f(a))), ... } (the true atoms)

Sentences P(a) and ∀x[P(x) → P(f(x))] are true, but P(b) is false.

Truth in Structures

4av

- The truth of a sentence S written in L under interpretation I is defined by:

S is a ground atom P(t1,...,tn): S is true iff val(I(P)(I(t1),...,I(tn))) = True

S =¬S1: S is true iff val(S1) = False

S = S1 op S2: S is true iff val(S1) op val(S2) is true (op in {∨, →, ∧, ↔})

S = ∀x(S1): S is true iff val(I(S1)(x/d)) is true for every d in D

S = ∃x(S1): S is true iff val(I(S1)(x/d) is true for some d in D

I(S1)(x/d) means d replaces occurrences of x in interpreted atoms in S1

A structure I for L is a *model* for a set of sentences S (written in L) if
for every sentence s in S, val(S) is true under I

If S has a model it is *satisfiable*. If S has no models S is *unsatisfiable*.

Exercise: what is the truth of ∃y∀x(P(x,y)) and ∀u∃v(P(u,v))
in the structure with Domain Integers
where I(P) = greater-than (ie P(x,y) means x >y)

Herbrand Interpretations

4avi

Some Definitions: Let L be a language for a set of sentences S

The *Herbrand Universe* HU of L is the set of terms using constants and function symbols in Sig(L). (It is assumed there is always at least one constant.)

The *Herbrand Base* HB of L is the set of ground atoms using terms from HU

An *Herbrand Interpretation* HI of L is a subset of the atoms in HB;
they are the set of true atoms in HB = {c: val(c) = true and c is in HB}

An *Herbrand model* of S is an Herbrand interpretation M of L
that forces val(S) =True for each sentence in S under M

If S has a Herbrand model we say S is *H-satisfiable*. If not S is *H-unsatisfiable*.

NOTE: If Sig(L) includes any function symbols then the Herbrand Universe is infinite.

There is assumed always one constant in HU, so HU ≠ ∅.

Herbrand Interpretations (Example)

4avii

Sig(L) = < {P,Q,R,S}, {f}, {a,b} >
S=Px ∨ Ry ∨ ¬Qxy, ¬Sz ∨ ¬Rz, Sa, ¬Pf(a) ∨ ¬Pf(b)

(Notation: Px is shorthand for P(x), Qxy is shorthand for Q(x,y), etc.)

- Herbrand Universe = {a,b,f(a),f(b),f(f(a)),f(f(b)), ...}
- Herbrand Base = {Pa,Pb, Pf(a), Pf(b), ... Sa, Sb, Sf(a), Sf(b), ..., Ra, Rb, Rf(a), ..., Qab, Qaa, Qbb,Qba, Qf(a)a, ...}
- One Herbrand interpretation is
  {Sa, all Q atoms except Qaa and Qbb, Pa, Pb, Pf(a), Pf(b), P(f(f(a)),..., }

meaning
val(Pa) = val(Pb) = ... = True, val(Sa) = True, val(any other S atom)=False,
val(Qaa) = val(Qbb) = False, val(any other Q atom) = False
and val(any R atom) = False

- This is **not** a Herbrand *model* of S because val(¬Pf(a) ∨ ¬Pf(b)) = False.
- The HI = { "S" atoms } **is** a model of S.

Soundness and Completeness Properties

4bi

Given a first order language L and sets of sentences A and B written in L:
A |= B - (A logically implies B) means that
whenever a structure M (of L) is a model of A,
then M is a model of at least one sentence in B.
Usually B is a single sentence, so M must be a model of B in this case

There are an infinite number of structures to check using this definition

So how else could we check A |=B (or equivalently that A,¬B |=⊥)?

We'll convert A and ¬B into clauses and use resolution to derive [ ]

S $\vdash_{res}$ [] - means [] can be deduced from S using resolution

How do we know we get the correct answers?

The two relations |= and $\vdash_{res}$ are equivalent,
as expressed by the *Soundness* and *Completeness* properties:

Soundness - if A $\vdash_{res}$ B then A |= B

Completeness - if A |= B then A $\vdash_{res}$ B

Soundness and Completeness of Resolution

4bii

Let C0 be a set of clauses. Let ⇒* denote "*yields by ≥1 resolution or factor steps*"

Soundness of Resolution: if C0 ⇒* [ ] then C0 |= ⊥ (or C0 has no models)

The idea used to show soundness is this:
show that for each resolution step
if the parent clauses are true in a structure then so is the resolvent,
and by transitivity that if the initial clauses are true so are all the resolvents.
Then, if a resolvent is clearly not true in any structure,
we conclude the given clauses are not true either.

We need (1) **Useful Theorem (*)** which states that
S has a Herbrand model iff S has some model
≡ S has no Herbrand models iff S has no models
Hence to show C0 |= ⊥ it is sufficient to show C0 $\models_H$ ⊥

and (2) a single resolution or factoring step is sound with respect to H-models:
if S ⇒ R then S $\models_H$ R (where R is a resolvent or factor from S)

where S $\models_H$ R holds iff
for every M, if M is a H-model of S then M is a H-model of R.
(Details and proofs of (1) and (2) are in Slides A1b in Appendix 1).

Proving Soundness of Resolution

4biii

Soundness of Resolution: if C0 ⇒* [ ] then C0 $\models \bot$ (or C0 has no models)

Using (1) and (2) from Slide 4bii we argue as follows

for a refutation C0⇒C0+C1⇒C0+C1+C2 ⇒ ... ⇒C0+...+[ ]:

by (2) (==> reads as implies)

C0 has H-model ==> C0+C1 has H-model ==> C0+C1+C2 has H-model ==> ... ==> C0+C1+...+Cn has H- model ≡

C0+C1+...+Cn has no H-model ==> C0+C1+...+Cn-1 has no H-model ==> ... ==> C0+C1 has no H-model ==> C0 has no H-model

Now suppose that [] is a resolvent (Cn say);
since [] has no models, C0+C1+...+Cn has no H-models,
hence .... C0 has no H-models.

Hence by (1) C0 has no models at all

Completeness of Resolution: if C0 $\models \bot$ then C0 ⇒* [ ]

Completeness is considered on Slides 5

The General Case

4di

We want to show that resolution can be used to show unsatisfiability of any set of sentences. Recall that conversion to clauses used Skolemisation (Step 3):

3. *Skolemise* - existential-type quantifiers are removed and bound variable occurrences of x in ∃xS are replaced by *Skolem constants* or *Skolem functions*

All non-*Skolemisation* steps in the conversion to clausal form are equivalences.
But, although Skolemised(S) $\models$ S, it is **not** true that S $\models$ Skolemised(S).

eg f(a) |= ∃x.f(x) – if f(a) is true then there is an x (namely a) s.t. f(x) is true.
But ∃x.f(x) does not imply f(a). Whatever x makes f(x) true need not be a.

However, it **is** true that Skolem(S) is unsatisfiable iff S is unsatisfiable. **(**)**
And this is what we need. (See Slides A1c in Appendix 1 for proof.)

In general:
To show Data |=Conclusion we convert {Data, ¬Conclusion} to clauses C.
Then
Data |=Conclusion iff {Data,¬Conclusion} is unsatisfiable (by definition)
iff C has no models (by (**) above)
iff C has no H-models **(by Useful Theorem (*) on 4bii)**
iff C⇒* [ ] (by Soundenss and Completeness of resolution)

Summary of Slides 4

4dii

1. Herbrand interpretations are first order structures which use a fixed mapping of terms in the Language to the structure. In particular, terms (constants or functional terms such as f(a)) map to the themselves.

2. Any set of sentences S has a model iff S has a Herbrand model.

3. Resolution is sound and complete: Derivation of [ ] from a set of clausesS by resolution and factoring implies that S|=⊥ , and if S|=⊥ then there is a resolution (and factoring) derivation of [ ] from S.

4. Soundness of resolution depends on the soundness of a single resolution or factoring step: if S⇒R then $S\models_H R$ and hence $S\models_H S+\{R\}$.

5. Resolution can be used to show S |= C, for arbitrary sentences S and C by first converting S and ¬C to clauses (Clauses(S+¬C)) and then showing that Clauses(S+¬C) ⇒* [ ] by resolution. By the Soundness of Resolution this means that $Clauses(S+\neg C)\models_H \bot$ and by (**) on 4di that S+¬C are unsatisfiable and hence that S |= C.