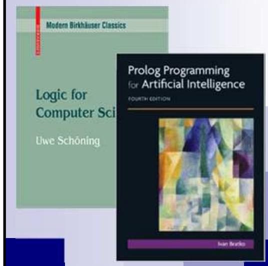

FLORIDA INTERNATIONAL UNIVERSITY

First Order Logic (III)

Dr. Antonio L. Bajuelos

School of Computing & Information Sciences

The SAT problem

- Remember that the **main goal** of our course is:
  - the search for an algorithmic test for **satisfiability** or **validity** of formulas.
  - We discussed in classes that the **SAT problem** for formulas in **Propositional Logic** is **NP-Complete**.

and in FOL?

  - **The SAT problem: Given a FOL formula F, the** satisfiability problem is concerned with the following question: Is there some interpretation I = $(\mathcal{U},\mathcal{A})$, such that I $\models$ F?
  - The **validty problem**: Given a FOL formula F, the validity problem is concerned with the following question: Is it the case that for all interpretations I = $(\mathcal{U},\mathcal{A})$, I $\models$ F?

2

No Model of Finite size...

Fact!

- There exist formulas in **FOL** which are **satisfiable**, but have **no models of finite size** (i.e. with a finite universe).
- **Example.** Consider a formula

$$
\begin{aligned}
F \;=\; & \forall x P(x, f(x)) \\
& \wedge \forall y \neg P(y, y) \\
& \wedge \forall u \forall v \forall w ((P(u, v) \wedge P(v, w)) \rightarrow P(u, w)).
\end{aligned}
$$

- This formula *F* is **satisfiable**, because it has for example the following model:

$$
\begin{aligned}
U_{\mathcal{A}} &= \{0, 1, 2, 3, \ldots\} = \mathbb{N} \\
P^{\mathcal{A}} &= \{(m, n) \mid m < n\}, \\
f^{\mathcal{A}}(n) &= n + 1.
\end{aligned}
$$

- But this formula does not possess a **finite model**.

Undecidability

Important points!

- In **computability theory**, a function is called **computable** (or a problem is called **decidable**) if there is an <u>abstract mathematical machine</u> (**Turing-machine**) which,
  - **started** with an input which is in the function domain,
  - **halts** after a <u>finite number of steps</u> and
  - **outputs** the correct function value (answers correctly "yes" or "no", according to the problem definition).
- If no such machine exists, then the function (problem) is called **non-computable** (**undecidable**).

Undecidability

- The **main problem**:

  Instance: A formula F in **First Order Logic**.

  Question: Is F valid?

- **Theorem (Church):**

  The **validity problem** for formulas of **FOL** is **undecidable.**

  (*see proof in the text book*)

- **Corollary:**

  The **satisfiability** problem of **FOL.**

  Instance: A formula F of **FOL**.

  Question: Is F **satisfiable**?

  is **undecidable**.

Undecidability

- **Corollary:**

  The **satisfiability** problem of **FOL**.

  Instance: A formula F of **FOL**.

  Question: Is F **satisfiable**?

  is **undecidable**.

- **Proof:**
  - A formula F is valid if and only if $\neg$F is unsatisfiable. Therefore, the hypothetical existence of a decision algorithm for the satisfiability problem leads to a decision algorithm for the validity problem, and we know by Church's theorem that such an algorithm does not exist. ■

Herbrand's theory

Remember that:

- One problem with dealing with formulas in **FOL** is that the definition of structures allows arbitrary sets as possible universes (domain).
- But we know that the problem of determining whether a given formula has a model or not is undecidable. So, we cannot expect to devise a decision algorithm.

Good news!

- Instead, we try to fix a special **domain** (called a **Herbrand* universe**) such that the formula, F, is **unsatisfiable** iff it is false under all the interpretations over this domain.

*__Jacques Herbrand__ (Paris, 1908 - 1931). Introduced the notion of recursive function and worked with John von Neumann and Emmy Noether. Died falling from a mountain in the Alps while climbing.

Herbrand universe

- The starting point of the **Herbrand's theory** are **closed formulas**, i.e. formulas without occurrences of free variables, which are in **Skolem normal form**.

Definition (Herbrand universe):

- The **Herbrand universe, D(F)**, of a closed formula **F** in **Skolem form** is:
  - The set of all **variable-free terms** that can be built from the components of F.
  - In the special case that F does not contain a constant, we first choose an arbitrary constant, say a, and then build up the variable-free terms.

Herbrand universe (cont…)

Definition (Herbrand universe):

- **D(F)** (or **H**) is defined inductively as follows:

  1. Every constant occurring in F is in D(F). If F does not contain a constant, then a is in D(F).

  2. For every k-ary function symbol f that occurs in F, and for all terms $t_1, t_2, \ldots, t_k$ already in D(F), the term $f(t_1, t_2, \ldots, t_k)$ is in D(F).

- **Example #1:**

  Consider the formula:

$$F = \forall x \forall y \forall z P(x, f(y), g(z, x))$$

  then the **Herbrand universe** is:

$$D(F) = \{a, f(a), g(a,a), f(g(a,a)), f(f(a)), g(a, f(a)), g(f(a), a), g(f(a), f(a)), \ldots\}$$

9

Herbrand universe (cont…)

Definition (Herbrand universe):

- **D(F)** (or **H**) is defined inductively as follows:

  1. Every constant occurring in F is in D(F). If F does not contain a constant, then a is in D(F).

  2. For every k-ary function symbol f that occurs in F, and for all terms $t_1, t_2, \ldots, t_k$ already in D(F), the term $f(t_2, t_2, \ldots, t_k)$ is in D(F).

- **Example #2:**

  Consider the formula:

$$G = \forall x \forall y Q(c, f(x), h(y, b))$$

  then the **Herbrand universe** is:

$$D(G) = \{c, b, f(c), f(b), h(c,c), h(c,b), h(b,c), h(b,b), f(f(c)), f(f(b)), f(h(c,c)), f(h(c,b)), f(h(b,c)), \ldots\}$$

10

Herbrand universe (cont…)

- **Example #3:**

$$F = \{P(a), \neg P(x) \vee P(f(x)), Q(x)\}$$
$$H_0(F) = \{a\}$$
$$H_1(F) = \{a, f(a)\}$$
$$H_2(F) = \{a, f(a), f(f(a))\}$$
$$\vdots$$
$$H_\infty(F) = \{a, f(a), f(f(a)), f(f(f(a))), \ldots\}$$

Herbrand universe (cont…)

- **Compact* definition:**

$$\begin{aligned}
Const(F) &= \text{set of constant symbols in } F \\
Fun(F) &= \text{set of function symbols in } F \\
H_0 &= \begin{cases} Const(F) & \text{if } Const(F) \neq \emptyset \\ \{a\} & \text{if } Const(F) = \emptyset \end{cases} \\
H_{i+1} &= \{f(t_1, .., t_n) \mid t_j \in (H_0 \cup .. \cup H_i),\ f/n \in Fun(F)\} \\
H(F) &= H_0 \cup .. \cup H_i \cup .. \qquad \text{is the Herbrand universe}
\end{aligned}$$

- **Examples:**

$F = \{p(x),\ q(y)\}$
- $H_0 = \{a\}$
- $H_1 = H_2 = .. = \emptyset$
- $H(F) = \{a\}$

$F = \{p(x, a),\ q(y) \vee \neg r(b, f(x))\}$
- $H_0 = \{a, b\}$
- $H_1 = \{f(a), f(b)\}$
- $H_2 = \{f(f(a)), f(f(b))\}$
- ...
- $H(F) = \{a, b, f(a), f(b), f(f(a)), f(f(b)), f(f(f(a))), f(f(f(b))), ..\} = \{f^n(a), f^n(b)\}_{n \geq 0}$

*http://ocw.upm.es/ciencia-de-la-computacion-e-inteligencia-artificial/computational-logic/contenidos/04interpretation.pdf

Herbrand base of *F*

- **ground atom**: an atom which is obtained by applying a predicate symbol of *F* to terms from the **Herbrand universe** of *F*
- the **Herbrand base** of *F* is the set of all the possible ground atoms of *F*
- **Example #1:**

$$F = \{P(a), \neg P(x) \vee P(f(x)), Q(x)\}$$

$$H_0 = \{a\}$$

$$H_1 = \{a, f(a)\}$$

$$H_2 = \{a, f(a), f(f(a))\}$$

$$\vdots$$

Let, $\mathrm{C} = Q(x)$

Here, $Q(a)$ and $Q(f(f(a)))$ are both ground atoms of C

$$\mathrm{HB} = \{P(a), Q(a), P(f(a)), Q(f(a)), \ldots\}$$

13

Herbrand base of *F* (cont…)

- **ground atom**: an atom which is obtained by applying a predicate symbol of *F* to terms from the **Herbrand universe** of *F*
- the **Herbrand base** of *F* is the set of all the possible ground atoms of *F*

Other examples:

- **F = {P(x), Q(y)}**
  - H(F) = {a}
  - HB(F) = {P(a), Q(a)}
- **F = {P(a), Q(y) ∨ ¬P(f(x))}**
  - H(F) = {a, f(a), f(f(a)),…}
  - HB(F) = {P(a), P(f(a)), P(f(f(a))),…,Q(a), Q(f(a)), Q(f(f(a))),…}
- **F = {P(a), Q(y) ∨ ¬R(b, f(x))}**
  - H(F) = {a, b, f(a), f(b), f(f(a)), f(f(b)),…}
  - HB(F) = ∪({{P(t), Q(t), R(t,t')}| t, t' ∈ H(F)})

14

Herbrand interpretations

- For a set of clauses S with its **Herbrand universe H**, we define I as an **H-Interpretation** if:
  - I maps all constants in S to themselves
  - An n-place function f is assigned a function that maps $(h_1,\ldots,h_n)$ to $f(h_1,\ldots,h_n)$ where $h_1,\ldots,h_n$ are elements in H or simply stated as $I = \{m_1, m_2, \ldots, m_n, \ldots\}$ where $m_j = A_j$ or $\neg A_j$ (i.e. A_j is set to true or false) and

$$A = \{A_1, A_2, \ldots, A_n, \ldots\}$$

Herbrand interpretations (cont…)

Example:

$$Let, S = \{P(x) \vee Q(x), R(f(y))\}$$

$$\Rightarrow \text{Herbrand Universe } H = H_\infty = \{a, f(a), f(f(a)), \ldots\}$$

$$\Rightarrow \text{Herbrand Base } HB \text{ is given by}$$

$$HB = \{P(a), P(f(a)), P(f(f(a))), \ldots, Q(a), \ldots, R(a), \ldots\}$$

$$\Rightarrow \text{Some Herbrand Interpretations are}$$

$$I_1 = \{P(a), P(f(a)), P(f(f(a))), \ldots, Q(a), \ldots, R(a), \ldots\}$$

$$I_2 = \{\neg P(a), \neg P(f(a)), \neg P(f(f(a))), \ldots, \neg Q(a), \ldots, \neg R(a), \ldots\}$$

$$I_3 = \{\neg P(a), \neg P(f(a)), \neg P(f(f(a))), \ldots, Q(a), \ldots, R(a), \ldots\}$$

Herbrand interpretations (cont…)

Example:

- F = {p(x), q(y)}

  H(F) = {a}, HB(F) = {p(a), q(a)}

  there are 4 possible **Herbrand interpretations**:

  I_1 = {p(a), q(a)}; I_2 = {p(a), ¬q(a)};

  I_3 = {¬p(a), q(a)}; I_4 = {¬p(a), ¬q(a)}

- F = {p(a), q(y) ∨ ¬p(f (x))}

  H(F) = $\{f^n(a)\}_{n\geq 0}$, HB(F) = ∪({{p(t), q(t)}| t ∈ H(F)})

  there are an <u>infinite number</u> of **Herbrand interpretations**

  I_1 = ∪({{p(t), q(t)} | t ∈ H(F)});

  I_2 = {p(a)} ∪ {¬p(t) | t ∈ H(F) \ {a}} ∪ {q(t) | t ∈ H(F)}

  I_3 = {p(t) | t ∈ H(F)} ∪ {¬q(t) | t ∈ H(F)} . . .

17

Herbrand interpretations (cont…)

- A **ground instance** of a clause is a formula, in clause form, which results from replacing the variables of the clause by terms from its **Herbrand universe**
- by means of an **Herbrand interpretation**, it is possible to give a truth value to a formula starting from the truth value of its **ground instances**
- **Example:** F = {p(a), q(b) ∨ ¬p(x)}

  H(F) = {a, b}

  HB(F) = {p(a), p(b), q(a), q(b)}

  I_H = {p(a), ¬p(b), q(a), ¬q(b)}

  - the **first clause is true** since its only instance, p(a), is true in I_H
  - the **second clause is false** since one instance, q(b) ∨ ¬p(b), is true in I_H , but the other, q(b) ∨ ¬p(a), is false
  - since F is the conjunction of both clauses, F is false for I_H

18

Herbrand's theory – main theorem

Theorem:

- A formula F is **unsatisfiable** iff it is **false for all** its **Herbrand interpretations**

Theorem:

- Let F be a closed formula in **Skolem form**. Then F is **satisfiable** if and only if F has a **Herbrand model**.

Example:

- F = {p(x), q(y)}, H(F) = {a}, HB(F) = {p(a), q(a)}
- There are 4 **Herbrand interpretations**
- I_1 = {p(a), q(a)}; I_2 = {p(a), ¬q(a)};
- I_3 = {¬p(a), q(a)}; I_4 = {¬p(a), ¬q(a)}
- **Ground instances**: {p(a), q(a)}
- I_1 is a model since it verifies both instances
- I_2, I_3 and I_4 are **countermodels** since they falsify at least one instance Therefore, **F is satisfiable**

19