Semi-Decidability of First-Order Logic

Marco Piastra

Decidability and automation of L_{FO}

- L_{FO} is <u>not</u> decidable

  No Turing machine can tell whether $\Gamma \models \varphi$

  Are there any hopes for automating the calculus?

- L_{FO} is ***semi-decidable*** **(Herbrand, 1930)**

  A Turing machine can tell (in *finite* time) that

  $$\Gamma \models \varphi$$

  ... but <u>not</u> that

  $$\Gamma \not\models \varphi$$

  In other words, the above Turing machine, when facing the problem "$\Gamma \models \varphi$?" :

  1) it will terminate with success if $\Gamma \models \varphi$
  2) it <u>*might*</u> diverge if $\Gamma \not\models \varphi$

Herbrand's System

Given a universal sentence of the form:

$\forall x_1 \forall x_2 \ldots \forall x_n \varphi$ (where φ does not contain quantifiers)

the ***Herbrand's System*** is the set (possibly *infinite*) of *ground* wffs generated by replacing the variables

$\varphi[x_1/t_1, x_2/t_2 \ldots x_n/t_n]$

A <u>*ground*</u> term or wff does not contain variables

with all possible combinations of *ground* terms $<t_1, t_2 \ldots t_n>$ of the *signature* Σ

Examples:

$\mathrm{H}(\forall x\, P(x) \rightarrow Q(x))) = \{P(f(a)) \rightarrow Q(f(a)), P(g(a, b)) \rightarrow Q(g(a, b)), \ldots \}$

$\mathrm{H}(\forall x\, \forall y\, R(x, y)) = \{R(f(a), f(a)), R(g(a, b), f(a)), R(f(a), g(a, b)), \ldots \}$

- ***Herbrand's System*** of a theory

Given a theory Φ of universal sentences, the Herbrand's system $\mathrm{H}(\Phi)$ is the union of all Herbrand's systems of the sentences in Φ

Example:

$\Phi = \{\varphi, \psi, \chi\}$

$\mathrm{H}(\Phi) = \mathrm{H}(\psi) \cup \mathrm{H}(\varphi) \cup \mathrm{H}(\chi)$

Herbrand's Theorem

- **Herbrand's Theorem**

  Given a theory of universal sentences Φ,
  $H(\Phi)$ has a model iff Φ has a model

  ... but what is the utility of that?

  $H(\Phi)$ may well be infinite even when Φ is finite,
  Furthermore, the theorem applies only to sets of universal sentences…

Prenex normal form (PNF)

Any wff φ can be transformed into an equivalent formula of the form

$$Q_1 x_1 Q_2 x_2 \ldots Q_n x_n \psi$$ (ψ is called the ***matrix***)

where Q_i is either $\forall$ or $\exists$ and ψ does not contain quantifiers

Equivalences:

$\models (\neg\forall x\, \varphi) \leftrightarrow (\exists x\, \neg\varphi)$ $\models (\neg\exists x\, \varphi) \leftrightarrow (\forall x\, \neg\varphi)$

$\models ((\forall x\, \varphi) \land \psi) \leftrightarrow (\forall x\, (\varphi \land \psi))$ $\models ((\exists x\, \varphi) \land \psi) \leftrightarrow (\exists x\, (\varphi \land \psi))$

$\models ((\forall x\, \varphi) \lor \psi) \leftrightarrow (\forall x\, (\varphi \lor \psi))$ $\models ((\exists x\, \varphi) \lor \psi) \leftrightarrow (\exists x\, (\varphi \lor \psi))$

$\models (\varphi \rightarrow (\forall x\, \psi)) \leftrightarrow (\forall x\, (\varphi \rightarrow \psi))$ $\models (\varphi \rightarrow (\exists x\, \psi)) \leftrightarrow (\exists x\, (\varphi \rightarrow \psi))$

However:

$\models ((\forall x\, \varphi) \rightarrow \psi) \leftrightarrow (\exists x\, (\varphi \rightarrow \psi))$ $\models ((\exists x\, \varphi) \rightarrow \psi) \leftrightarrow (\forall x\, (\varphi \rightarrow \psi))$

Caution: *variables MUST be renamed, when required, in order to avoid clashes*

Examples: $\exists y\, (P(y) \rightarrow \forall x\, P(x))$

$\exists y \forall x\, (P(y) \rightarrow P(x))$ (PNF, using $(\varphi \rightarrow (\forall x\, \psi)) \leftrightarrow (\forall x\, (\varphi \rightarrow \psi))$)

$\exists y\, (\forall x\, P(x) \rightarrow P(y))$

$\exists y \exists x\, (P(x) \rightarrow P(y))$ (PNF, using $((\forall x\, \varphi) \rightarrow \psi) \leftrightarrow (\exists x\, (\varphi \rightarrow \psi))$)

$\forall x \exists y\, (Q(x,y) \rightarrow P(y)) \land \neg\forall x\, P(x)$

$\forall x \exists y\, (Q(x,y) \rightarrow P(y)) \land \exists x\, \neg P(x)$ (Using $(\neg\forall x\, \varphi) \leftrightarrow (\exists x\, \neg\varphi)$)

$\forall x \exists y\, (Q(x,y) \rightarrow P(y)) \land \exists z\, \neg P(z)$ (*substitution* $[x/z]$)

$\forall x \exists y \exists z\, ((Q(x,y) \rightarrow P(y)) \land \neg P(z))$ (PNF)

Skolemization

In a sentence in PNF, existential quantifiers can be eliminated by extending the *signature* Σ of the *language*

Consider a sentence in PNF $Q_1 x_1 Q_2 x_2 \ldots Q_n x_n \psi$

From left to right, for each $Q_i x_i$ of type $\exists x_i$:

- Apply to ψ the *substitution* $[x_i / k(x_1, \ldots, x_j)]$ where k is a new function and $x_1, \ldots, x_j$ are the variables of j the universal quantifiers that come *before* $\exists x_i$ (k is an individual constant if $j = 0$)
- $\exists x_i$ is simply removed

Examples:

$\exists y\, \forall x\, (P(y) \to P(x))$

$\forall x\, (P(k) \to P(x))$ (k Skolem's constant)

$\forall x\, \exists y\, \exists z\, ((Q(x,y) \to P(y)) \wedge \neg P(z))$

$\forall x\, ((Q(x, k(x)) \to P(k(x))) \wedge \neg P(m(x)))$ ($k/1$ and $m/1$ Skolem's functions)

- Theorem

For any sentence φ,

φ has a model iff $sko(\varphi)$ (i.e. Skolemization of φ) has a model

Semi-decidability of L_{PO}

- Corollary of Herbrand's theorem

  These three statements are equivalent:

  - $\Gamma \models \varphi$
  - $\Gamma \cup \{\neg\varphi\}$ is not satisfiable (= it has no model)
  - There exist a ***finite*** subset
    of $\mathrm{H}(sko(\Gamma \cup \{\neg\varphi\}))$ (= Herbrand's system of the Skolemitazion of $\Gamma \cup \{\neg\varphi\}$)
    that is ***inconsistent***

  Therefore:

  When $\Gamma \models \varphi$, a procedure that generates the finite *subsets* of $\mathrm{H}(sko(\Gamma \cup \{\neg\varphi\}))$
  will certainly discover a contradiction (*in finite time*)