

# Intelligenza Artificiale II

## Semi-decidibilità di $L_{PO}$

Marco Piastra

## Decidibilità ed automazione di $L_{PO}$

- Indecidibilità di  $L_{PO}$

Non esiste un algoritmo (di valore generale) in grado di stabilire se  $\Gamma \models \varphi$

Al contrario del caso proposizionale, in  $L_{PO}$  non si possono verificare direttamente tutte le possibili interpretazioni

*Qual è quindi la speranza di avere un calcolo automatico?*

- In realtà,  $L_{PO}$  è **semi-decidibile** (Herbrand, 1930)

E` possibile stabilire (in tempo finito) se

$$\Gamma \models \varphi$$

... ma non se

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

- In altri termini, possono esistere metodi che:

Di fronte al problema “ $\Gamma \models \varphi$  ?”

a) Terminano con successo se  $\Gamma \models \varphi$

b) Possono divergere (i.e. girare all'infinito) se  $\Gamma \not\models \varphi$

# Universo e base di Herbrand

## Termini e atomi di Herbrand

Dato un linguaggio  $L_{PO}$

Un **termine** di Herbrand è un termine chiuso (*ground term*, = che non contiene variabili)

Esempi:

$f(a), g(a,b), g(f(a),b), g(f(a),g(b,c)), g(f(a),g(f(b),c)), \dots$

Un **atomo** di Herbrand è una fbf *atomica* (*ground term*, = che non contiene variabili)

Esempi:

$P(f(a)), P(g(a,b)), Q(g(f(a),b), g(f(a),g(b,c))), \dots$

## Universo e base di Herbrand

L'**universo** di Herbrand è l'insieme di tutti i termini di Herbrand

Esempio:

$U_H \equiv \{f(a), g(a,b), g(f(a),b), g(f(a),g(b,c)), g(f(a),g(f(b),c)), \dots\}$

La **base** di Herbrand è l'insieme di tutti gli atomi di Herbrand

Esempio:

$B_H \equiv \{P(f(a)), P(g(a,b)), Q(g(f(a),b), g(f(a),g(b,c))), \dots\}$

# Modelli di Herbrand

- **Struttura di Herbrand per  $L_{PO}$**

Una struttura  $\langle \mathbf{U}_H, v_H \rangle$  tale che

$$\forall c \in \text{Cost}(L_{PO}), v_H(c) = c$$

$$\forall t \in \mathbf{U}_H, v_H(t) = t$$

- **Interpretazione  $v_H$  di Herbrand**

Un qualsiasi **sottoinsieme** della base di Herbrand  $B_H$

$$v_H \equiv \{P(a), P(f(b)), P(c), Q(a, g(b, c)), Q(b, c) \dots\} \quad (\text{solo formule atomiche chiuse})$$

$$v_H \subseteq B_H$$

- **Modello di Herbrand**

$$\varphi \in \text{Atom}(L_{PO}), \langle \mathbf{U}_H, v_H \rangle [s] \models \varphi \quad \text{sse} \quad \varphi \in v_H$$

$$\varphi \in \text{Atom}(L_{PO}), \langle \mathbf{U}_H, v_H \rangle [s] \models \neg \varphi \quad \text{sse} \quad \varphi \notin v_H$$

$$\langle \mathbf{U}_H, v_H \rangle [s] \models \neg \varphi \quad \text{sse} \quad \langle \mathbf{U}_H, v_H \rangle [s] \not\models \varphi$$

$$\langle \mathbf{U}_H, v_H \rangle [s] \models \varphi \rightarrow \psi \quad \text{sse non} \quad (\langle \mathbf{U}_H, v_H \rangle [s] \models \varphi \text{ e } \langle \mathbf{U}_H, v_H \rangle [s] \not\models \psi)$$

$$\langle \mathbf{U}_H, v_H \rangle [s] \models \forall x \varphi \quad \text{se per ogni } c \in \text{Cost}(L_{PO}) \text{ si ha } \langle \mathbf{U}_H, v_H \rangle [s](x:c) \models \varphi$$

# Sistemi di Herbrand

- **Sistema di Herbrand di un enunciato universale**

Dato un enunciato universale, della forma

$$\forall x_1 \forall x_2 \dots \forall x_n \varphi \quad (\varphi \text{ non contiene quantificatori})$$

il **Sistema di Herbrand** è l'insieme (anche infinito) di formule chiuse generato per sostituzione

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

con tutte le possibili combinazioni  $\langle t_1, t_2 \dots t_n \rangle$  di  $t_i \in U_H$

Esempi:

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

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

- **Sistema di Herbrand di una teoria**

Data una teoria  $\Sigma$  di enunciati universali

è l'unione  $H(\Sigma)$  di tutti i sistemi di Herbrand generati dagli enunciati  $\Sigma$

Esempio:

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

$$H(\Sigma) = H(\psi) \cup H(\varphi) \cup H(\chi)$$

# Teorema di Herbrand

- Teorema di Herbrand

Data una teoria di enunciati universali  $\Sigma$ ,  
 $H(\Sigma)$  ha un modello sse  $\Sigma$  ha un modello

... ma qual'è l'utilità?

$H(\Sigma)$  può essere infinito anche quando  $\Sigma$  è finito  
il teorema si applica solo agli enunciati universali

# Forma normale prenessa

## ▪ Forma normale prenessa FNP (i.e. tutti i quantificatori all'inizio)

Una fbf  $\varphi$  qualsiasi può essere trasformata in un enunciato equivalente

$$Q_1x_1Q_2x_2 \dots Q_nx_n \psi \quad (\psi \text{ è anche detta } \mathbf{matrice})$$

dove  $Q_i$  è  $\forall$  oppure  $\exists$  e  $\psi$  non contiene quantificatori

Si ottiene da  $\varphi$  eliminando tutte le negazioni  $\neg$  davanti ai quantificatori tramite la definizione  $\neg\forall x \equiv \exists x \neg$  e le equivalenze derivanti

$$(\alpha \rightarrow \forall x \beta) \equiv \forall x (\alpha \rightarrow \beta) \quad \text{dato che } (\alpha \rightarrow \forall x \beta) \equiv (\neg\alpha \vee \forall x \beta) \equiv \forall x (\neg\alpha \vee \beta)$$

$$(\forall x \alpha \rightarrow \beta) \equiv \exists x (\alpha \rightarrow \beta) \quad \text{dato che } (\forall x \alpha \rightarrow \beta) \equiv (\neg\forall x \alpha \vee \beta) \equiv (\exists x \neg\alpha \vee \beta) \equiv \exists x (\neg\alpha \vee \beta)$$

E' necessario *ridenominare* le variabili per evitare sovrapposizioni

Esempio:  $\exists y (P(y) \rightarrow \forall x P(x))$

$$\exists y \forall x (P(y) \rightarrow P(x)) \quad (\text{FNP, usando } (\varphi \rightarrow \forall x \psi) \leftrightarrow \forall x (\varphi \rightarrow \psi))$$

Esempio:  $\exists y (\forall x P(x) \rightarrow P(y))$

$$\exists y \exists x (P(x) \rightarrow P(y)) \quad (\text{FNP, usando } (\forall x \varphi \rightarrow \psi) \leftrightarrow \exists x (\varphi \rightarrow \psi))$$

Esempio:  $\forall x \exists y (Q(x,y) \rightarrow P(y)) \wedge \neg \forall x P(x)$

$$\forall x \exists y (Q(x,y) \rightarrow P(y)) \wedge \exists x \neg P(x) \quad (\text{definizione } \neg \forall x \varphi \equiv \exists x \neg \varphi)$$

$$\forall x \exists y (Q(x,y) \rightarrow P(y)) \wedge \exists z \neg P(z) \quad (\text{ridenominazione di } x \text{ in } z)$$

$$\forall x \exists y \exists z ((Q(x,y) \rightarrow P(y)) \wedge \neg P(z)) \quad (\text{FNP})$$

# Forma di Skolem

## Forma di Skolem

Si eliminano i quantificatori esistenziali in una forma normale prenessa sostituendo ciascuna variabile esistenzialmente quantificata con una (nuova) costante o una (nuova) funzione:

Si opera sui quantificatori della formula  $Q_1x_1Q_2x_2 \dots Q_nx_n \psi$  partendo da sinistra

Per ogni  $Q_ix_i$  della forma  $\exists x_i$ :

- si applica a  $\psi$  la sostituzione  $[x_i/k_i(x_1, \dots, x_{i-1})]$  (nuova funzione di Skolem) oppure  $[x_i/k_i]$  (nuova costante di Skolem), se  $i = 1$
- si elimina  $\exists x_i$  dalla formula

Esempi:

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

$$\forall x (P(k) \rightarrow P(x))$$

( $k$  costante di Skolem: si espande il linguaggio)

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

$$\forall x ((Q(x, k(x)) \rightarrow P(k(x))) \wedge \neg P(j(x))) \quad (k/1 \text{ e } j/1 \text{ funzioni di Skolem})$$

## Teorema

Per qualsiasi enunciato  $\varphi$ ,

$\varphi$  ha un modello sse  $sko(\varphi)$  (forma di Skolem di  $\varphi$ ) ha un modello



## Semi-decidibilità effettiva di $L_{PO}$

### ■ Corollario del teorema di Herbrand

Le seguenti affermazioni sono equivalenti

- $\Gamma \models \varphi$
- $\Gamma \cup \{\neg\varphi\}$  non è soddisfacibile (= non ha un modello) (= è inconsistente)
- Esiste un sottoinsieme **finito** di  $H(sko(\Gamma \cup \{\neg\varphi\}))$  (sistema di Herbrand della forma di Skolem) che è **contraddittorio**

Quindi:

Una procedura che enumera tutti i sottoinsiemi finiti di  $H(sko(\Gamma \cup \{\neg\varphi\}))$  prima o poi (*in un tempo finito*) trova una contraddizione, sse  $\Gamma \models \varphi$

Esempio:  $\forall x (P(f(x)) \rightarrow P(g(x))), \exists y \neg P(g(y)) \models \exists z \neg P(f(z))$

- 1:  $\{\forall x (P(f(x)) \rightarrow P(g(x))), \exists y \neg P(g(y)), \neg \exists z \neg P(f(z))\}$  ( $\Gamma \cup \{\neg\varphi\}$ )
- 2:  $\{\forall x (P(f(x)) \rightarrow P(g(x))), \exists y \neg P(g(y)), \forall z P(f(z))\}$  (definizione di  $\exists$ )
- 3:  $\{\forall x (P(f(x)) \rightarrow P(g(x))), \neg P(g(k)), \forall z P(f(z))\}$  ( $sko(\Gamma \cup \{\neg\varphi\})$ )
- 4:  $\{(P(f(k)) \rightarrow P(g(k))), \neg P(g(k)), P(f(k))\}$  ( $\in H(sko(\Gamma \cup \{\neg\varphi\}))$ ),  $[x/k, z/k]$ )
- 5:  $\{(\neg P(f(k)) \vee P(g(k))), \neg P(g(k)), P(f(k))\}$  (traduzione di  $\rightarrow$ )
- 6:  $\{\neg P(f(k)), P(f(k))\}$  (risoluzione di  $P(g(k)), \neg P(g(k))$ )  
(*contraddizione*)