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 <u>non</u> 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:

  $f(a)$, $g(a,b)$, $g(f(a),b)$, $g(f(a),g(b,c))$, $g(f(a),g(f(b),c))$, …

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

  Esempi:

  $P(f(a))$, $P(g(a,b))$, $Q(g(f(a),b), g(f(a),g(b,c)))$, …

- Universo e base di Herbrand

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

  Esempio:

  $\mathbf{U}_H \equiv \{f(a), g(a,b), g(f(a),b), g(f(a),g(b,c)), g(f(a),g(f(b),c)), \dots\}$

  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 $<\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}>$ tale che

  $\forall c \in \mathrm{Cost}(L_{PO}),\ v_{\mathrm{H}}(c) = c$

  $\forall t \in \mathbf{U}_{\mathrm{H}},\ v_{\mathrm{H}}(t) = t$

- Interpretazione v_{H} di Herbrand

  Un qualsiasi sottoinsieme della base di Herbrand B_{H}

  $v_{\mathrm{H}} \equiv \{P(a), P(f(b)), P(c), Q(a,g(b,c)), Q(b,c) \ldots \}$ (solo formule atomiche chiuse)

  $v_{\mathrm{H}} \subseteq \mathrm{B}_{\mathrm{H}}$

- Modello di Herbrand

  $\varphi \in \mathrm{Atom}(L_{PO}),\ <\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}> [s] \models \varphi$ sse $\varphi \in v_{\mathrm{H}}$

  $\varphi \in \mathrm{Atom}(L_{PO}),\ <\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}> [s] \models \neg\varphi$ sse $\varphi \notin v_{\mathrm{H}}$

  $<\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}> [s] \models \neg\varphi$ sse $<\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}> [s] \not\models \varphi$

  $<\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}> [s] \models \varphi \rightarrow \psi$ sse non $(<\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}> [s] \models \varphi$ e $<\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}> [s] \not\models \psi)$

  $<\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}> [s] \models \forall x\, \varphi$ se per ogni $c \in \mathrm{Cost}(L_{PO})$ si ha $<\mathbf{U}_{\mathrm{H}}, v_{\mathrm{H}}> [s](x{:}c) \models \varphi$

Sistemi di Herbrand

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

  Dato un enunciato universale, della forma

  $\forall x_1 \forall x_2 \ldots \forall x_n \varphi$ (φ non contiene quantificatori)

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

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

  con <u>tutte</u> le possibili combinazioni $<t_1, t_2 \ldots t_n>$ di $t_i \in \mathbf{U}_\mathrm{H}$

  Esempi:

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

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

- **Sistema** di Herbrand di una teoria

  Data una teoria Σ di enunciati universali
  è l'unione $\mathrm{H}(\Sigma)$ di tutti i sistemi di Herbrand generati dagli enunciati Σ

  Esempio:

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

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

Teorema di Herbrand

- **Teorema di Herbrand**

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

  ... ma qual'è l'utilità?
  $H(\Sigma)$ può essere infinito anche quando Σ 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 φ qualsiasi può essere trasformata in un enunciato equivalente

  $Q_1 x_1 Q_2 x_2 \ldots Q_n x_n \psi$ (ψ è anche detta ***matrice***)

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

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

  $(\alpha \to \forall x\, \beta) \equiv \forall x\, (\alpha \to \beta)$ dato che $(\alpha \to \forall x\, \beta) \equiv (\neg\alpha \vee \forall x\, \beta) \equiv \forall x\, (\neg\alpha \vee \beta)$

  $(\forall x\, \alpha \to \beta) \equiv \exists x\, (\alpha \to \beta)$ dato che $(\forall x\, \alpha \to \beta) \equiv (\neg\forall x\, \alpha \vee \beta) \equiv (\exists x\, \neg\alpha \vee \beta) \equiv \exists x\, (\neg\alpha \vee \beta)$

  E` necessario *ridenominare* le variabili per evitare sovrapposizioni

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

  $\exists y \forall x\, (P(y) \to P(x))$ (FNP, usando $(\varphi \to \forall x\, \psi) \leftrightarrow \forall x\, (\varphi \to \psi)$)

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

  $\exists y \exists x\, (P(x) \to P(y))$ (FNP, usando $(\forall x\, \varphi \to \psi) \leftrightarrow \exists x\, (\varphi \to \psi)$)

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

  $\forall x \exists y\, (Q(x,y) \to P(y)) \wedge \exists x\, \neg P(x)$ (definizione $\neg\forall x\, \varphi \equiv \exists x\, \neg\varphi$)

  $\forall x \exists y\, (Q(x,y) \to P(y)) \wedge \exists z\, \neg P(z)$ (*ridenominazione* di x in z)

  $\forall x \exists y \exists z\, ((Q(x,y) \to P(y)) \wedge \neg P(z))$ (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 $Q_1 x_1 Q_2 x_2 \ldots Q_n x_n \psi$ partendo da sinistra

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

  - si applica a ψ la sostituzione $[x_i/k_i(x_1, \ldots, x_{i-1})]$ (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)

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

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

- Teorema

  Per qualsiasi enunciato φ,

  φ ha un modello sse $sko(\varphi)$ (forma di Skolem di φ) 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 $\mathrm{H}(sko(\Gamma \cup \{\neg\varphi\}))$ (sistema di Herbrand della forma di Skolem) che è **contraddittorio**

  Quindi:

  Una procedura che enumera _tutti_ i sottoinsiemi finiti di $\mathrm{H}(sko(\Gamma \cup \{\neg\varphi\}))$ 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: $\{(P(f(k)) \rightarrow P(g(k))), \neg P(g(k)), P(f(k))\}$ $(\in \mathrm{H}(sko(\Gamma \cup \{\neg\varphi\})), [x/k, z/k])$
  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*)