Oltre la logica classica: ragionamento plausibile

Marco Piastra

Oltre la logica classica?

- Per logica **classica** si intende:
  - La logica predicativa del primo ordine L_{PO}
  - La logica proposizionale L_P (che è contenuta, in senso proprio, in L_{PO})

- Una logica **non classica** adotta regole diverse
- *Perchè?*
  - Per rappresentare altre forme di ragionamento
    - Non solo deduttivo ma anche *abduttivo* ed *induttivo* (*vedi oltre*)
    - Forme speciali, legate ad obiettivi specifici, come logiche modali o temporali
  - Per esigenze applicative
    - Frammenti (sottoinsiemi) di L_{PO}, più efficacemente automatizzabili (p.es. Jess)

Logiche e sistemi logici

In ambito teorico (p.es. in logica matematica) una **logica** è definita da:

a) Linguaggio formale
b) Semantica del linguaggio formale
c) Relazioni $\models$ (*conseguenza*) e $\vdash$ (*derivazione*)

- In intelligenza artificiale

E` utile vedere un **sistema logico** come un *agente ragionatore*

- Basato su una **logica** di riferimento (p.es. L_{PO})
- Adotta una determinata **strategia di calcolo** (p.es. *SLD depth-first*)
- Può avere risorse limitate (di tempo o memoria)

Si ha quindi il concetto di *derivabilità* in un **sistema logico**

Notazione: $\Gamma \vdash_{<SysLog>} \varphi$ dove $<SysLog>$ indica un **sistema logico** particolare

Esempi:

Strategia *SLD* (definita solo per le clausole di Horn) che sia anche *fair*

$$\Gamma \vdash_{LPO} \varphi \neq \Gamma \vdash_{SLD\ fair} \varphi \neq \Gamma \vdash_{SLD} \varphi$$

↑ Derivabilità generale in L_{PO}

Strategia *SLD* non necessariamente *fair* (e.g. *depth-first*)

In linea di principio, la strategia di calcolo di $<SysLog>$ può essere qualsiasi cosa:
p.es. $\Gamma \vdash_{NN} \varphi$ una rete neurale che stabilisce se φ è (NN) derivabile da Γ

Ragionamento plausibile (*defeasible reasoning*)

- In generale

Un ragionamento dove la **relazione** tra premesse e conseguenza
è <u>razionalmente plausibile</u> ma non necessariamente <u>corretta</u> (in senso logico-classico)

Notazione:
$\Gamma \mid\!\sim_{<SysLog>} \varphi$ indica che φ è una derivazione **plausibile** (con premesse Γ) in un sistema $<SysLog>$

Proprietà della relazione $\mid\!\sim_{<SysLog>}$

$\Gamma \mid\!\sim_{<SysLog>} \varphi \Rightarrow \Gamma \not\mid\!\sim_{<SysLog>} \neg\varphi$ (coerente)

$\Gamma \vdash_{<SysLog>} \varphi \Rightarrow \Gamma \mid\!\sim_{<SysLog>} \varphi$ (chiusa rispetto alla derivazione)

$\Gamma \mid\!\sim_{<SysLog>} \varphi \not\Rightarrow \Gamma \vdash_{<SysLog>} \varphi \; (\Rightarrow \Gamma \models \varphi)$ (non necessariamente corretta)

Molto frequente in pratica:

"L'orario ferroviario non riporta un treno per Milano alle 06:55,
quindi si assume che tale treno <u>non</u> esista"

In generale, un database contiene solo informazione positiva (p.es. i treni esistenti)

L'informazione negativa è ricavata 'per difetto'

Ragionamento plausibile (*defeasible reasoning*)

- Inferenza non monotona

  In generale, non vale la proprietà di monotonia

  $\Gamma \vdash_{<SysLog>} \varphi \nRightarrow \Gamma \cup \Delta \vdash_{<SysLog>} \varphi$

  L'arrivo di nuova informazione può infatti falsificare una precedente inferenza

  p.es. l'annuncio di un treno straordinario …

- Inferenza sistemica

  Nel caso del *modus ponens*, si ha uno schema di inferenza di valore generale

  $\varphi \to \psi, \varphi \vdash \psi$

  (E` sempre applicabile, non dipende dal contesto)

  In generale, le inferenze *plausibili* dipendono da un'intera teoria Γ

  "L'orario ferroviario non riporta un treno per Milano alle 06:55,
  quindi si assume che tale treno non esista"

  Vale a dire: $\Gamma \nvdash \varphi \quad \vdash_{<SysLog>} \neg\varphi$

  Tipicamente, la teoria Γ rappresenta *un sistema completo di conoscenze* (p.es. un database)

  Qualsiasi modifica di Γ, in generale, può cambiare l'insieme $\{\varphi \mid \Gamma \vdash_{<SysLog>} \varphi\}$

  Occorre controllare tutto il database (l'orario) per essere sicuri che non vi sia l'indicazione del treno

Closed-World Assumption (CWA)

$\{\Gamma \not\models \alpha\} \vdash_{CWA} \neg\alpha$ (α è un *atomo*)

Esempio:

$\Pi \equiv \{\{Filosofo(socrate)\}, \{Filosofo(platone)\}, \{Mortale(felix)\}\}$

I fatti del programma Π possono essere riscritti in L_{PO} come:

$\forall x\ ((x = socrate) \rightarrow Filosofo(x))$
$\forall x\ ((x = platone) \rightarrow Filosofo(x))$
$\forall x\ ((x = felix) \rightarrow Mortale(x))$

La *CWA* equivale al *completamento* (***plausibile***) di Π:

$\forall x\ ((x = felix) \leftrightarrow Mortale(x))$
$\forall x\ ((x = socrate \lor x = platone) \leftrightarrow Filosofo(x))$

Notare la doppia implicazione

Inferenza plausibile

In questo caso:

$\Pi \vdash_{CWA} \neg Mortale(socrate)$
$\Pi \vdash_{CWA} \neg Mortale(platone)$
$\Pi \vdash_{CWA} \neg Filosofo(felix)$

CWA e regole

Esempio:

$\Pi \equiv \{\{\neg Filosofo(x), Umano(x)\}, \{\neg Umano(y), Mortale(y)\},$
$\{Filosofo(socrate)\}, \{Filosofo(platone)\}, \{Mortale(felix)\}\}$

La riscrittura dei fatti del programma Π è identica al caso precedente:

$\forall x\, ((x = socrate) \rightarrow Filosofo(x))$
$\forall x\, ((x = platone) \rightarrow Filosofo(x))$
$\forall x\, ((x = felix) \rightarrow Mortale(x))$

Ma il *completamento* di Π deve tener conto delle regole:

$\forall x\, ((x = felix \vee x = socrate \vee x = platone) \leftrightarrow Mortale(x))$
$\forall x\, ((x = socrate \vee x = platone) \leftrightarrow Filosofo(x))$
$\forall x\, ((x = socrate \vee x = platone) \leftrightarrow Umano(x))$

Inferenza plausibile

In questo caso:

$\Pi \mid\!\sim_{CWA} \neg Umano(felix)$
$\Pi \mid\!\sim_{CWA} \neg Filosofo(felix)$

Negazione come fallimento (*Negation as Failure - NF*)

Non tutte le fbf sono traducibili in clausole di Horn

In particolare non lo è:

$(\neg A \land B) \rightarrow C$

$\neg(\neg A \land B) \lor C$

$A \lor \neg B \lor C$ (si ottiene una clausola con due letterali positivi)

Si legga il connettivo $\neg$ nella premessa della regola come '*non è noto che*'

Esempio:

$(\neg A \land B) \rightarrow C$

Se A non è noto e B è vero, allora C

Salvo *"diversa comunicazione"* e *"oggi è venerdì"*, allora *"c'è lezione di IA2"*

- La Negazione come Fallimento (NF)

In un sistema logico <*SysLog*>

Si legge il connettivo $\neg$ nella premessa come '*non è dimostrabile da* <*SysLog*>'

Si tratta di un'inferenza *plausibile*

Risoluzione SLDNF

- La NF nel sistema logico della risoluzione SLD

  $(\neg A \wedge B) \rightarrow C$ nella notazione SLDNF si traduce come

  $\backslash + \neg A \vee \neg B \vee C$ (il simbolo \+¬ indica la *NF*)

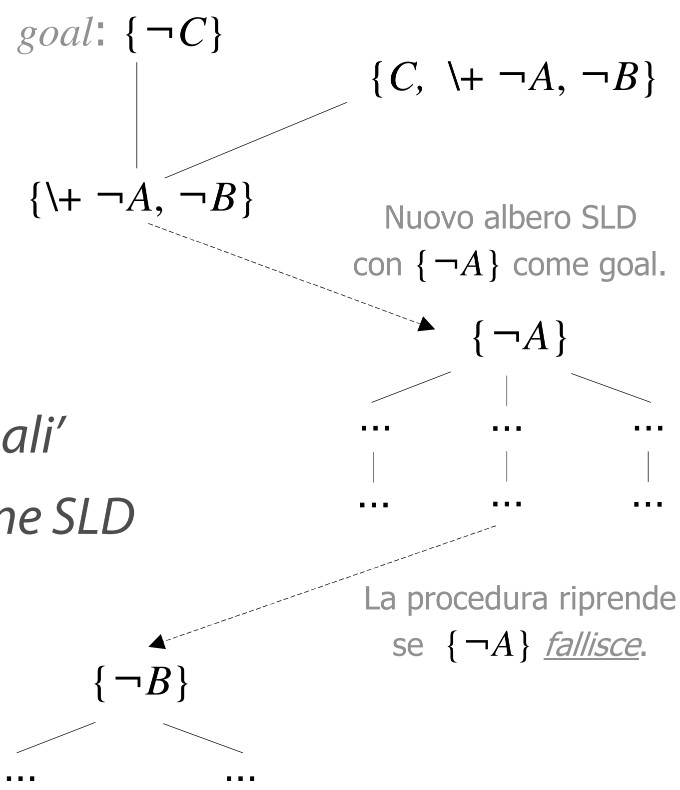
- Metodo

  Le premesse negative (= precedute da \+¬)
  sono considerate come
  goal la cui derivazione SLD deve *fallire*

  Il metodo SLDNF procede come SLD per i goal 'normali'

  Incontrando un $\backslash + \neg \varphi$ *, si apre una nuova derivazione SLD*

  Il goal $\backslash + \neg \varphi$ *ha successo se* $\neg \varphi$ <u>*fallisce*</u>



Risoluzione SLDNF e completamento

- Completamento
  1) Dato un insieme di clausole di Horn Γ, si aggregano le regole in modo che ciascun atomo compaia al massimo in una implicazione

     Esempio: $\Gamma \equiv \{\{C\}, \{B, \neg F\}, \{B, \neg E\}, \{B, \neg D\}\}$

     Si può riscrivere: $\Gamma \equiv \{C, F \to B, E \to B, D \to B\}$

     In forma aggregata: $\Gamma' \equiv \{C, D \vee E \vee F \to B\}$
  2) Per ogni atomo φ non fattuale si aggiunge $false \to \varphi$ ($false$ = *contraddizione*)

     Nell'esempio: $\Gamma'' \equiv \{C, D \vee E \vee F \to B, false \to D, false \to E, false \to F\}$
  3) Si sostituisce l'implicazione $\to$ con l'equivalenza $\leftrightarrow$

     $Comp(\Gamma'') \equiv \{C, D \vee E \vee F \leftrightarrow B, false \leftrightarrow D, false \leftrightarrow E, false \leftrightarrow F\}$ (completamento)

- *Correttezza* di SLDNF (Clark, 1974)

  Se il goal SLDNF φ ha successo per Γ allora $Comp(\Gamma) \models \varphi$

  Esempio: $\Gamma \equiv \{\{C\}, \{B, \neg F\}, \{B, \neg E\}, \{B, \neg D\}\}$

  Il goal $\neg B$ ha successo SLDNF perchè $\neg B$ *fallisce* in SLD

  In effetti: $\{C, (D \vee E \vee F) \leftrightarrow B, false \leftrightarrow D, false \leftrightarrow E, false \leftrightarrow F\} \models \neg B$

Risoluzione SLDNF

- Non è completa

  $Comp(\Gamma) \models \neg\varphi \;\not\Rightarrow\; \Gamma \vdash_{SLDNF} \neg\varphi$

  Ciò non dipende dalle *selection function*, che sia fair oppure no

  Tuttavia: $Comp(\Gamma) \models \neg P$

- Non è logica classica

  Non vale la proprietà di *monotonia sintattica*

  $\Gamma \vdash_{SLDNF} \neg\varphi \;\not\Rightarrow\; \Gamma \cup \Delta \vdash_{SLDNF} \neg\varphi$

  Esempio:

  $\{C, D \vee E \vee F \to B\} \vdash_{SLDNF} \neg B$

  $\{C, D \vee E \vee F \to B\} \cup \{D\} \nvdash_{SLDNF} \neg B$

  Aggiungendo informazione, il risultato della NF può cambiare

Relazione tra CWA, NF e SLDNF

- *Closed-World Assumption* (CWA)

  $\{\Gamma \not\models \alpha\} \vdash_{CWA} \neg\alpha$ (α è un *atomo*)

  Notare che $\Gamma \not\models \alpha$ non è decidibile in L_{PO}, e quindi nemmeno la relazione $\vdash_{CWA}$

- SLDNF *fair*

  $\{\alpha \in FF_{SLDfair}(\Gamma)\} \vdash_{SLDNF\ fair} \neg\alpha$

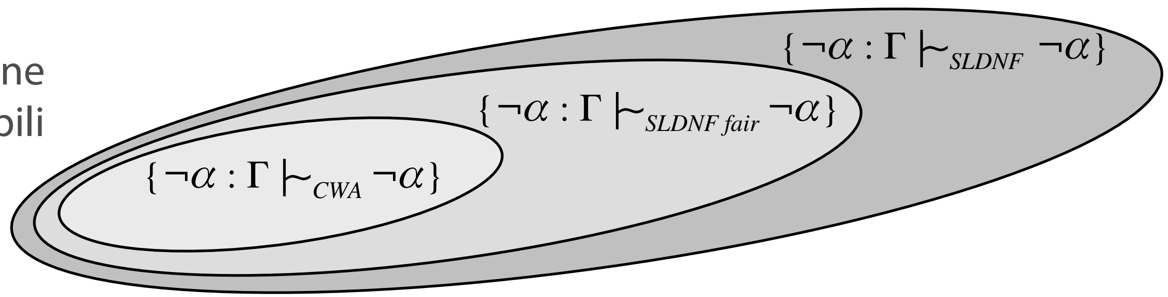
  $FF_{SLDfair}(\Gamma)$ è il set di goal atomici per cui la procedura SLD *fair* fallisce senza andare in loop *(Finite Failure)*

- SLDNF

  $\{\alpha \in FF_{SLD}(\Gamma)\} \vdash_{SLDNF} \neg\alpha$

  Come sopra, ma senza assumere che la SLD sia *fair*

Relazioni di inclusione
tra insiemi di letterali derivabili



Default Logic (Reiter, 1980)

- Descrizione intuitiva

Regole di default

Sono regole specifiche, espresse con una notazione 'ad hoc'

Le premesse sono di due tipi: *vere* (*giustificate*) e solo *plausibili*

$$\frac{\langle giustificazione \rangle : \langle plausibilità \rangle}{\langle default \rangle}$$

schema informale

Una regola di default

$$\rho: \ \frac{\alpha : \beta_1, \beta_2, \ldots, \beta_n}{\gamma}$$

notazione formale

è applicabile ad una *teoria* (=un insieme di fbf) B
sse $\mathrm{B} \vdash \alpha$ e $\mathrm{B} \nvdash \neg\beta_1, \mathrm{B} \nvdash \neg\beta_2, \ldots, \mathrm{B} \nvdash \neg\beta_n$

Allora si ha

$$(\mathrm{B}, \{\rho\}) \vdash_{default} \gamma$$

Default Logic (Reiter, 1980)

Esempi:

$B \equiv \{\{\neg Filosofo(x), Umano(x)\}, \{\neg Umano(y), Mortale(y)\},$
$\{Filosofo(socrate)\}, \{Filosofo(platone)\}, \{Mortale(felix)\}\}$

Considerando la regola di default:

$$\rho_1: \frac{Mortale(x) : Umano(x)}{Umano(x)}$$

si ha che $(B,\{\rho_1\}) \mid\!\sim_{default} Umano(felix)$

In alternativa, considerando:

$$\rho_2: \frac{Mortale(x) : \neg Gatto(x), Umano(x)}{Umano(x)}$$

si ha che $(B \cup \{Gatto(felix)\},\{\rho_1\}) \mid\!\sim_{default} Umano(felix)$
ma $(B \cup \{Gatto(felix)\}, \{\rho_2\}) \not\mid\!\sim_{default} Umano(felix)$

La *closed world assumption* (CWA) equivale ad adottare lo *schema* di regola di default:

$$\rho_{CWA}: \frac{: \neg\varphi}{\neg\varphi}$$

in questo caso si ha quindi che:

$(B,\{\rho_{CWA}\}) \mid\!\sim_{default} \neg Filosofo(felix)$
$(B,\{\rho_{CWA}\}) \mid\!\sim_{default} \neg Umano(felix)$

Default Logic (Reiter, 1980)

Teorie

Una *teoria* in Default Logic è formata da una coppia (Γ, Δ):

Γ rappresenta l'informazione certa (in senso logico-classico)

Δ contiene le regole di ragionamento plausibile (regole di default)

Estensioni

Un'estensione è una teoria (=insieme di *fbf*) che risulta chiuso rispetto alla deduzione classica e un dato insieme di regole di default.

Data una teoria (Γ, Δ), Σ è un'*estensione* di (Γ, Δ)

a) se è deduttivamente chiusa, (in senso classico) $\Sigma \supseteq \{\alpha \mid \Sigma \vdash \alpha\}$

b) ed è il risultato dell'applicazione incrementale di tutte le regole in Δ secondo un determinato ordine (un *processo*)

Il risultato può infatti cambiare a seconda dell'ordine di applicazione delle regole di default…

Default Logic (Reiter, 1980)

Esempio

Il risultato può infatti cambiare a seconda dell'ordine di applicazione delle regole di default…

La teoria di default:

$$(\{\ Republican(Nixon), Quaker(Nixon)\ \},\ \{\ \frac{Republican(x) : \neg Pacifist(x)}{\neg Pacifist(x)}\quad \frac{Quaker(x) : Pacifist(x)}{Pacifist(x)}\ \})$$

Ha due estensioni, a seconda dell'ordine di applicazione delle regole:

Σ_1: $\{\ Republican(Nixon), Quaker(Nixon), \neg Pacifist(x)\}$

Σ_2: $\{\ Republican(Nixon), Quaker(Nixon), Pacifist(x)\}$

Default Logic (Reiter, 1980)

Proprietà

Ciascuna *teoria* in Default Logic (Γ, Δ) può avere un numero qualsiasi di estensioni

Due forme di conseguenza per le fbf ottenute per default

Forma scettica

$(\Gamma, \Delta) \approx_{default} \gamma$ sse γ appartiene a <u>tutte</u> le estensioni di (Γ, Δ)

Forma credula

$(\Gamma, \Delta) \approx_{default} \gamma$ sse γ appartiene ad <u>almeno una</u> estensione di (Γ, Δ)

In entrambi i casi, nel campo della logica del primo ordine,
il problema $(\Gamma, \Delta) \approx_{default} \gamma$ non è decidibile

Nemmeno semi-decidibile

Forme di inferenza (C. S. Peirce)

- Nuovi schemi di ragionamento

Logica deduttiva (inferenza *corretta*)

Si derivano (solo) le conseguenze logiche delle premesse

"Tutti i fagioli del sacco sono bianchi,
questa manciata di fagioli viene dal sacco"
"I fagioli sono bianchi"

Modus Ponens

$$\frac{\varphi \to \psi \qquad \varphi}{\psi}$$

Logica induttiva (inferenza *plausibile*)

Da associazioni di fatti, si derivano regole

"Questa manciata di fagioli viene dal sacco,
i fagioli sono bianchi"
"Tutti i fagioli del sacco sono bianchi"

$$\frac{\varphi \qquad \psi}{\varphi \to \psi}$$

Logica abduttiva (inferenza *plausibile*)

Da regole ed effetti si derivano le cause

"Tutti i fagioli del sacco sono bianchi,
i fagioli sono bianchi"
"Questa manciata di fagioli viene dal sacco"

$$\frac{\varphi \to \psi \qquad \psi}{\varphi}$$

Attenzione! Non è il Modus Ponens

Abduzioni come ipotesi esplicative

- La logica di base è la logica **classica**

  E` invece diverso il **tipo** di ragionamento

  e quindi il **tipo** di calcolo utilizzato

- In generale, in un ragionamento abduttivo:

  Un *modello* (o descrizione astratta)
  rappresentato da una teoria K

  Un insieme di *osservazioni* specifiche
  rappresentate da un insieme di fbf Σ

  In generale, $\mathrm{K} \not\models \Sigma$
  (le osservazioni possibili non sono conseguenza logica di K)

  Si cerca è un'ipotesi Δ tale per cui

  $$\mathrm{K} \cup \Delta \models \Sigma$$

  Intuitivamente, Δ descrive un'ipotesi che *spiega* Σ

Esempio: "La macchina non parte"

- Modello (K)

  κ_1: $batteriaScarica \rightarrow (\neg funzionanoLuci \wedge \neg funzionaAutoradio \wedge \neg motorinoGira)$

  κ_2: $motorinoGuasto \rightarrow \neg motorinoGira$

  κ_3: $\neg motorinoGira \rightarrow \neg macchinaParte$

  κ_4: $serbatoioVuoto \rightarrow (indicatoreAZero \wedge \neg macchinaParte)$

- Osservazioni (Σ)

  σ_1: $\neg macchinaParte$

- Possibili *ipotesi* (Δ)

  δ_1: $batteriaScarica$ infatti: $\{\kappa_1, \kappa_3\} \cup \{\delta_1\} \models \sigma_1$

  δ_2: $motorinoGuasto$ infatti: $\{\kappa_2, \kappa_3\} \cup \{\delta_2\} \models \sigma_1$

  δ_3: $serbatoioVuoto$ infatti: $\{\kappa_4\} \cup \{\delta_3\} \models \sigma_1$

Razionalità delle ipotesi

- Ipotesi *plausibili*
  - Le ipotesi Δ devono essere *consistenti* con la teoria e le osservazioni
    - $K \cup \Delta \cup \Sigma$ deve essere soddisfacibile
  - Le ipotesi Δ devono spiegare *tutte* le osservazioni Σ
    - Alcune ipotesi, tuttavia, implicano anche altre osservazioni:
    - $batteriaScarica \rightarrow (\neg funzionanoLuci \wedge \neg funzionaAutoradio \wedge \neg motorinoGira)$
  - Le ipotesi Δ devono essere *minimali*
    - Non deve esistere un $\Delta^* \subset \Delta$ tale per cui $K \cup \Delta^* \vdash \Sigma$
    - Le ipotesi devono limitarsi ai soli elementi indispensabili per spiegare Σ
- Rilevanza
  - Notare che:
    $K \cup \{\neg macchinaParte\} \models \neg macchinaParte$
    è un'ipotesi è plausibile ma inutile: non ha valore esplicativo
  - Le ipotesi devono risalire ad elementi che abbiano un qualche valore *causale*
    - La cui definizione spesso dipende dall'ambito della teoria

Scelta tra ipotesi

- Ipotesi multiple possono coesistere

  Le due ipotesi $\{serbatoioVuoto\}$ e $\{batteriaScarica\}$ sono consistenti (possono coesistere)

  $K \cup \{serbatoioVuoto, batteriaScarica\} \models \neg macchinaParte$

- Altre ipotesi possono essere alternative

  Le due ipotesi $\{serbatoioVuoto, funzionaAutoradio\}$ e $\{batteriaScarica\}$ sono plausibili ma mutuamente esclusive

  $K \cup \{serbatoioVuoto, funzionaAutoradio, batteriaScarica\}$ è inconsistente

- Strategie di scelta tra ipotesi

  Spesso è utile ridurre il numero delle ipotesi (p.es. quando si cerca un rimedio)

  Acquisizione di nuove osservazioni

  Partendo dalle ipotesi $\{serbatoioVuoto\}$, $\{batteriaScarica\}$ e $\{motorino\ Guasto\}$ l'acquisizione dei valori di *funzionaAutoradio*, *motorinoGira* e *indicatoreAZero* permette di scegliere (purchè tali fatti siano *osservabili*)

  Criteri particolari

  Scelta basata sul *costo* delle osservazioni

  Scelta basata sul *rischio* associato alle ipotesi

Diverse forme di ragionamento abduttivo

La teoria K rappresenta l'informazione che l'agente usa per *spiegare* le osservazioni
La valenza informale della teoria K è fondamentale per le applicazioni pratiche

Fault Recognition ("La macchina non parte")

La teoria K descrive il rapporto causa-effetto delle possibili *anomalie* di un sistema
Le osservazioni Σ descrivono le anomalie effettivamente osservate

Ipotesi Δ: $\mathrm{K} \cup \Delta \models \Sigma$
Ciascuna ipotesi Δ descrive le possibili cause delle anomalie

Schema-based reasoning (SBR)

La teoria K descrive il *comportamento atteso* di un sistema
Le osservazioni Σ rappresentano le anomalie rispetto al comportamento atteso

Le osservazioni sono incompatibili con la teoria: $\mathrm{K} \cup \Sigma$ è *insodddisfacibile*
Ipotesi Δ: $(\mathrm{K} - \Delta) \models \Sigma$
Ciascuna ipotesi Δ descrive le deviazioni rispetto al comportamento atteso

Case-based reasoning (CBR)

Il mondo è regolare: problemi simili hanno spiegazioni simili

Posto che: $\mathrm{K} \not\models \Sigma$
Ipotesi Δ': $(\mathrm{K} - \Delta) \cup \Delta' \models \Sigma$
Ciascuna ipotesi Δ' descrive gli *adattamenti* delle informazioni al caso specifico