Plausible Reasoning

Marco Piastra

Beyond classical logic?

- For *classical* logic it usually intended:
  - First-order logic L_{FO}
  - Propositional logic L_P (which is contained in L_{FO})

- A **non-classical** logic adopts different rules
- *What for?*
  - Representing other forms of reasoning
    - Not just *deduction* but also *abduction* and *induction* (*see after*)
    - Specialized reasoning, e.g. about time or other modalities like belief, intentions etc.
  - For practical applications
    - Subsets of L_{FO}, that are either more efficient or focused on a specific purpose (e.g. Prolog)

Logics and logical systems

Theoretically, a **logic** includes:

a) Formal language

b) Formal semantics of the language

c) Relations $\models$ (*entailment*) e $\vdash$ (*derivation*)

- In the realm of artificial intelligence

A **logical system** is a *reasoning agent* (not necessarily human)

- It is based on the a **logic** of reference (e.g. L_{FO})
- It makes use of a **computation strategy** (e.g. *SLD depth-first*)
- It may have *limited resources* (e.g. time or memory or both)

This leads to the idea of *derivability* in a **logical system**

Notation: $\Gamma \vdash_{<SysLog>} \varphi$ where <*SysLog*> describes a particular logica system

Example:

SLD strategy (just for Horn clauses) that is also *fair*

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

↑ General derivability in L_{FO}

A generic *SLD* strategy (i.e. not necessarily *fair*)

In the line of principle, the computation strategy of <*SysLog*> can be anything:
e.g. $\Gamma \vdash_{NN} \varphi$ might ne a neural network that says whether φ is (*NN*)-derivable from Γ

Defeasible reasoning

A reasoning process where the **relation** between formulae
is <u>rationally plausible</u> yet not necessarily <u>correct</u> (in the classical logical sense)

Notation:
$\Gamma \mid\!\sim_{<SysLog>} \varphi$ says that φ is a **plausible** derivation from Γ in $<SysLog>$

Properties of $\mid\!\sim_{<SysLog>}$

$\Gamma \mid\!\sim_{<SysLog>} \varphi \Rightarrow \Gamma \not\mid\!\sim_{<SysLog>} \neg\varphi$	(coherence)
$\Gamma \vdash_{<SysLog>} \varphi \Rightarrow \Gamma \mid\!\sim_{<SysLog>} \varphi$	(compatibility with derivation)
$\Gamma \mid\!\sim_{<SysLog>} \varphi \not\Rightarrow \Gamma \vdash_{<SysLog>} \varphi \; (\Rightarrow \Gamma \models \varphi)$	(not necessarily correct)

Occurs very often in practice:

"The train schedule does not report a train to Milano at 06:55,
therefore we assume that such a train does <u>not</u> exist"

Most databases contain positive information only
Negative facts are often derived 'by default'

Defeasible reasoning

- Inference in *defeasible reasoning* is

Non-monotonic

$$\Gamma \mathrel{|\!\sim}_{<SysLog>} \varphi \;\not\Rightarrow\; \Gamma \cup \Delta \mathrel{|\!\sim}_{<SysLog>} \varphi$$

The arrival of new pieces of information may falsify inferences that used to be justified

e.g. an extra train to Milano at 06:55 is announced …

Systemic

In classical logic, the soundness of all inferences schema depend only on the few formulae involved.

e.g. $\varphi \to \psi, \varphi \vdash \psi$

In defeasible reasoning, inferences are justified by an entire theory Γ

One must check the entire database: $\Gamma \nvdash \varphi \quad \mathrel{|\!\sim}_{<SysLog>} \neg\varphi$

Closed-World Assumption (CWA)

$\{\Gamma \not\models \alpha\} \vdash_{CWA} \neg\alpha$ (α is an *atom*)

Example (a program):

$\Pi \equiv \{\{Philosopher(socrates)\}, \{Philosopher(plato)\}, \{Cat(felix)\}\}$

The program Π can be rewritten in L_{FO} as:

$\forall x\ ((x = socrates) \rightarrow Philosopher\ (x))$
$\forall x\ ((x = plato) \rightarrow Philosopher\ (x))$
$\forall x\ ((x = felix) \rightarrow Cat(x))$

The *Closed-World Assumption* (*CWA)* means completing (i.e. extending) the program Π:

$\forall x\ ((x = felix) \leftrightarrow Cat(x))$
$\forall x\ ((x = socrates \lor x = plato) \leftrightarrow Philosopher\ (x))$ *Notice the double implication*

Then these plausible inferences become sound:

$\Pi \vdash_{CWA} \neg Cat(socrates)$
$\Pi \vdash_{CWA} \neg Cat(plato)$
$\Pi \vdash_{CWA} \neg Philosopher\ (felix)$

Inference and reasoning (according to C. S. Peirce, 1870 c.a.)

- Different types of reasoning

  <u>Deductive</u> inference (sound)

  *Derive only what is justified in terms of **entailment***

  "All beans in this bag are white"
  "This handful of beans comes form the bag"
  "This is a handful of white beans"

$$\frac{\forall x\, \varphi(x) \rightarrow \psi(x) \quad \varphi(a)}{\psi(a)}$$

  <u>Inductive</u> inference (plausible)

  From repeated occurrences, derive rules

  "This handful of beans comes form the bag"
  "This is a handful of white beans"
  "All beans in this bag are white"

$$\frac{\psi(a) \quad \varphi(a)}{\forall x\, \varphi(x) \rightarrow \psi(x)}$$

  <u>Abductive</u> inference (plausible)

  From rules and outcomes, derive premises

  "All beans in this bag are white"
  "This is a handful of white beans"
  "This handful of beans comes form the bag"

$$\frac{\forall x\, \varphi(x) \rightarrow \psi(x) \quad \psi(a)}{\varphi(a)}$$

Abductive inferences: explanatory hypotheses

The basic theory is still that of classical logic

What changes is the way reasoning (and hence calculus) is performed

- Abductive reasoning, in general:

A ***model*** (or abstract definition of some kind)
represented by a logical theory K

A set of specific ***observations***
represented by a set of wffs Σ

In general: $K \not\models \Sigma$

(specific observations are not *entailed* by the model)

The problem is finding *hypotheses* Δ (i.e sets of wffs) such that

$$K \cup \Delta \models \Sigma$$

Intuitively, a set Δ describes an hypothesis that *explains* the observations Σ

Example: "The car does not start"

- Model (K)

  κ_1: *dischargedBattery* $\rightarrow$ ($\neg$*lightsOn* $\wedge$ $\neg$*radioOn* $\wedge$ $\neg$*selfStarterRuns*)

  κ_2: *selfStarterBroken* $\rightarrow$ $\neg$ *selfStarterRuns*

  κ_3: $\neg$*selfStarterRuns* $\rightarrow$ $\neg$*engineStarts*

  κ_4: *voidTank* $\rightarrow$ (*gasGaugeZero* $\wedge$ $\neg$*engineStarts*)

- Observation (Σ)

  σ_1: $\neg$*engineStarts*

- Plausible *causes* (Δ)

  δ_1: *dischargedBattery* since: $\{\kappa_1, \kappa_3\} \cup \{\delta_1\} \models \sigma_1$

  δ_2: *selfStarterBroken* since: $\{\kappa_2, \kappa_3\} \cup \{\delta_2\} \models \sigma_1$

  δ_3: *voidTank* since: $\{\kappa_4\} \cup \{\delta_3\} \models \sigma_1$

Rationality of hypotheses

- P*lausible*

  $K \cup \Delta \cup \Sigma$ must be *satisfiable*

- *Minimal*

  There must not be a subset $\Delta^* \subset \Delta$ such that $K \cup \Delta^* \models \Sigma$

- *Relevant*

  $K \cup \{\neg \mathit{engineStarts}\} \models \neg \mathit{engineStarts}$
  is both plausible and minimal but offers no explication
  (abductive reasoning is about the *causes*, in some sense)