VU Einführung in Wissensbasierte Systeme

WS 2010/11

Hans Tompits

Institut für Informationssysteme
Arbeitsbereich Wissensbasierte Systeme

www.kr.tuwien.ac.at

5. Nonmonotonic Reasoning

5.1 Introduction

Classical logic vs. commonsense reasoning

- Classical logic deals with the analysis of *truth* and *valid arguments*.
- A typical valid reasoning pattern is like the following:

  All men are mortal.
  Clark Kent is a man.
  Therefore, Clark Kent is mortal.

  Symbolised by classical first-order logic:

  $$\forall x(Man(x) \rightarrow Mortal(x)), Man(Clark\_Kent) \models Mortal(Clark\_Kent)$$

Classical logic vs. commonsense reasoning (ctd.)

- A valid argument remains valid even in the presence of new information:

  All men are mortal.
  Clark Kent is a man.
  Clark Kent is an alien from the planet Krypton.
  Therefore, Clark Kent is mortal.

- Even if we add *inconsistent* information to the premisses, the conclusion is still derivable:

  All men are mortal.
  Clark Kent is a man.
  Clark Kent is an alien from the planet Krypton.
  No alien is a man.
  Therefore, Clark Kent is mortal.

- In general, classical logic satisfies the *monotonicity principle*:
  - if $S \models A$ and $S \subseteq S'$, then $S' \models A$.

Classical logic vs. commonsense reasoning (ctd.)

- On the other hand, *human commonsense reasoning* deals with less strict reasoning patterns.
- A typical argument in commonsense reasoning is the following:

  Birds typically fly.
  Tweety is a bird.
  Therefore, Tweety flies.

Classical logic vs. commonsense reasoning (ctd.)

- Here, the conclusion is drawn in the *absence of information to the contrary*:
  - "Birds typically fly" means "given no information to the contrary, a bird flies".
  - Since all we know about Tweety is that it is a bird, we conclude that Tweety flies.

- A conclusion inferred this way is *plausible*, or *rational*, but may have to be *retracted* given more specific information:
  - if we later learn that Tweety is a penguin, it is no longer rational that Tweety flies since penguins do not fly.

  ➥ "Tweety flies" is no longer asserted!

Classical logic vs. commonsense reasoning (ctd.)

- To summarise: human commonsense reasoning involves a flexible form of reasoning
  - conclusions are drawn in the presence of *incomplete information*
  - they may have to be *retracted* given new and more accurate information ("jumping to conclusions")
  - assumptions are tentative, subject to revision
  - typically, current information is considered *the only relevant one* for a particular problem.
    Example: "He has not told me that he is his brother. So, I assume that he is not."

- Commonsense reasoning is *nonmonotonic*, violating the monotonicity principle.

Classical logic vs. commonsense reasoning (ctd.)

➤ As a consequence, classical logic is not adequate to model human commonsense reasoning

- classical logic is monotonic, satisfying the monotonicity principle, disallowing the revision of conclusions
- it presupposes *complete information* about a domain under consideration
- it only makes implicit knowledge explicit (correctness of classical logic)

➥ Other formalisms are necessary to formalise rational conclusions!

Formalisms for commonsense reasoning

- Different approaches for dealing with incomplete and uncertain information have been proposed:
  - quantitative methods:
    - using probability theory ($\rightarrow$ not discussed in this course)
  - qualitative methods:
    - nonmonotonic logics
- Most nonmonotonic logics have been defined in the early 1980s.
  - closed-world assumption (CWA)
  - circumscription
  - modal nonmonotonic logics
  - default logic

$\Longrightarrow$ here, we deal with the CWA and default logic

- Important computational approach for nonmonotonic reasoning:
  - logic programming under the answer-set semantics ($\rightarrow$ LU EWBS and VL Logikorientierte Programmierung)

Some notation

- A *theory* is a set of closed formulas.
- The *deductive closure* of a theory T is given by

$$Cn(T) = \{\varphi \mid T \vdash \varphi \text{ and } \varphi \text{ is closed}\},$$

  where $\vdash$ is the derivability relation of classical first-order logic.

- Some properties:
  - $T \subseteq Cn(T)$; ("inflationaryness")
  - $Cn(T) = Cn(Cn(T))$ ("idempotency");
  - $T \subseteq T'$ implies $Cn(T) \subseteq Cn(T')$ ("monotonicity");
  - $Cn(\emptyset)$ is the set of all valid formulas
    - thus, in particular, $Cn(\emptyset) \neq \emptyset$;
  - $Cn(T)$ is the set of all formulas iff T is inconsistent.
  - If $\varphi \in Cn(T)$, then $Cn(T \cup \{\varphi\}) = Cn(T)$.

5.2 Closed-world assumption

The closed-world assumption (CWA)

Method for an efficient representation of negative facts.

- Introduced by Raymond Reiter in 1978.

Observation: Number of negative facts is often much higher than number of positive facts.

Example: library database

Assumption: 1000 readers; 10.000 books;
each reader may borrow at most 5 books.

$\Longrightarrow$ not sensible to store 9.995.000$\sim$10.000.000 data entries of books per reader which are not borrowed!

Basic idea of CWA

- Store only positive information.
- Assumption:
  - every positive fact which is not derivable is assumed to be false!
  - Example: train timetable
    - ↪ Trains which are not explicitly mentioned in the timetable are assumed to be non-existing.
- CWA is nonmonotonic:
  - After adding new facts, some negative facts may no longer be derivable.

CWA: Formal definition

- Given: theory T (= set of closed formulas).
- Ground atoms: closed atomic formulas (i.e., variable-free atoms).
  - Examples: $p(a)$, $q(f(b, g(a)))$, ...
  - Here and henceforth, constants are denoted by $a, b, c, \ldots$ and variables by $x, y, z, \ldots$.
- Closed-world assumption (CWA) of T:

$$\begin{aligned} T_{asm} &= \{\neg P \mid P \text{ ground atom, } T \nvdash P\}; \\ \mathrm{CWA}(T) &= \{\varphi \mid T \cup T_{asm} \vdash \varphi,\ \varphi \text{ closed}\} = Cn(T \cup T_{asm}). \end{aligned}$$

  - Intuition: $\mathrm{CWA}(T)$ is the logical closure of all assumptions (explicit and implicit ones).

Some properties

Definition: A theory T is complete iff, for each ground atom P, $P \in T$ or $\neg P \in T$ holds.

Example: $T = \{p(a),\ (p(a) \rightarrow q(a)),\ p(b)\}$.

- T is not complete: $q(a), \neg q(a) \notin T$; $q(b), \neg q(b) \notin T$.
- But: $q(a) \in \mathrm{CWA}(T)$, since $T \vdash q(a)$;
  $\neg q(b) \in \mathrm{CWA}(T)$, since $T \nvdash q(b)$.

$\mathrm{CWA}(T)$ is complete $\Longrightarrow$ *completes* T by adding negative facts.

Some properties (ctd.)

Problem: Satisfiability (consistency) of the CWA?

Example: $T = \{p(a) \lor p(b)\}$.

- ➤ $T \nvdash p(a)$, $T \nvdash p(b)$.
- ➥ $\neg p(a) \land \neg p(b) \in \mathrm{CWA}(T)$;
- ➥ inconsistent with $p(a) \lor p(b)$.

Theorem: Let T be a consistent theory. Then:
$\mathrm{CWA}(T)$ is inconsistent $\iff$ there are ground atoms $A_1, \ldots, A_n$ such that $T \models A_1 \lor \ldots \lor A_n$, but $T \not\models A_i$, for all $i = 1, \ldots, n$.

Note: Result depends on the chosen language!

Some properties (ctd.)

Example: $T = \{\forall x(p(x) \vee q(x)),\ p(a),\ q(b)\}$

- Assumption: language contains only constants a and b $\Longrightarrow$ $\mathrm{CWA}(T)$ is consistent.
- Additional constant c $\Longrightarrow$ $\mathrm{CWA}(T)$ is inconsistent:
  $(T \vdash p(c) \vee q(c)$ and $T \not\vdash p(c),\ T \not\vdash q(c))$.

Theorem: Let T be a set of *definite Horn clauses*. Then, $\mathrm{CWA}(T)$ is consistent.

- Recall: A definite Horn clause is a clause having *precisely one positive literal*.
- Example: $\neg p \vee \neg q \vee s$ is a definite Horn clause; $\neg p \vee \neg q$ is *not* a definite Horn clause.

Generalisations of the CWA

- CWA relative to a predicate symbol $q \Longrightarrow$ application of the CWA only to predicates containing the predicate symbol q.

  Formally:

$$T^q_{asm} = \{\neg P \mid P \text{ ground atom with predicate symbol } q,\ T \not\vdash P\};$$
$$\mathrm{CWA}^q(T) = \{\varphi \mid T \cup T^q_{asm} \vdash \varphi,\ \varphi \text{ closed}\} = Cn(T \cup T^q_{asm}).$$

- CWA relative to a set $\{q_1, \ldots, q_n\}$ of predicate symbols $\Longrightarrow$ similarly defined!

- Example: Let T be the following set of formulas:

$$\forall x(q(x) \rightarrow p(x))$$
$$q(a)$$
$$r(b) \vee p(b)$$

  - CWA relative to p yields $\neg p(b) \in \mathrm{CWA}^p(T)$, and thus $r(b) \in \mathrm{CWA}^p(T)$.
  - Unrestricted CWA yields $\neg p(b) \wedge \neg r(b) \in \mathrm{CWA}(T) \Longrightarrow$ inconsistent with T!

5.3 Default logic

5.3.1 General considerations and basic definitions

Default logic

- Method for dealing with incomplete information and exceptions.
- Represent facts which hold *typically* by means of special inference rules (default rules).
- Default rules extend classical inference rules by additional *consistency conditions*.
- Default logic was introduced by Reiter in 1980.

Default rules

- A default rule (or simply a default), δ, is an inference rule of the form

$$\frac{\varphi : \psi_1, \ldots, \psi_n}{\chi}$$

- Intuitive meaning:
  - if φ is known and every ψ_i *can be consistently assumed* (i.e., $\neg\psi_i$ is *not* derivable), then infer χ.

- Notation:
  - $\varphi = pre(\delta)$: prerequisite;
  - $\{\psi_1, \ldots, \psi_n\} = just(\delta)$: justifications;
  - $\chi = cons(\delta)$: consequent.

Default rules: Examples

- Rules with exceptions:

$$\frac{bird(X)\ :\ can\_fly(X)}{can\_fly(X)}$$

  - Unless known otherwise, a bird can fly.
- Rules which hold in general, usually, or typically:

$$\frac{go\_to\_work\ :\ take\_bus}{take\_bus}$$

  - Typically, I take the bus when I go to work.
- Rules which hold unless the contrary is explicitly known:

$$\frac{accused\ :\ innocent}{innocent}$$

  - The accused is innocent unless proven otherwise.

Default theory

- In default logic, knowledge about the world is represented in terms of a default theory $T = (W, \Delta)$:

  W: set of closed formulas of first-order logic (*certain knowledge, premisses*);

  Δ: set of defaults (*plausible inferences*).

- Basic idea of default reasoning:
  - apply the defaults in Δ to the facts in W to derive plausible inferences from certain knowledge;
  - apply the defaults to the extended knowledge until no new knowledge is generated $\rightarrow$ *extension E of T*.

Extensions of default theories

An extension E should have the following properties:

- E is a *theory*, i.e., a set of *closed classical formulas*;
- E contains the facts: $W \subseteq E$.
- E is *deductively closed*, i.e., $Cn(E) = E$, since we want to derive more knowledge than by classical means.
- E is closed under applications of defaults, i.e., if $\delta = \frac{\varphi : \psi_1, \ldots, \psi_n}{\chi} \in \Delta$ is applicable to E, then $\chi \in E$, where:

$$\boxed{\delta \text{ is applicable to } E \text{ iff } \varphi \in E \text{ and } \neg\psi_1 \notin E, \ldots, \neg\psi_n \notin E}$$

Extensions of default theories (ctd.)

We have to address two issues:

1. How can extensions be defined formally?

   Applicability of defaults:

   - default prerequisite $\varphi \in E$;
   - default justification $\neg\psi_1 \notin E, \ldots, \neg\psi_n \notin E$ $\quad\longrightarrow$ difficult!

2. How can extensions be computed (efficiently)?

Example: Tweety

- Consider the following default theory $T = (W, \Delta)$:

$$\begin{aligned} W &= \{Bird(Tweety)\}; \\ \Delta &= \left\{ \delta = \frac{Bird(x) \;:\; Flies(x)}{Flies(x)} \right\}. \end{aligned}$$

- An extension E of T should contain $Bird(Tweety)$ and $Flies(Tweety)$, *and all formulas which can be classically derived from them*.
  $\Longrightarrow$ It should hold that $E = Cn(\{Bird(Tweety), Flies(Tweety)\})$.

Example: Tweety (ctd.)

➤ Consider now $T' = (W', \Delta)$, where

$$W' = W \cup \{Penguin(Tweety), \forall x(Penguin(x) \rightarrow \neg Flies(x))\}.$$

➥ Extension E' of T' should contain W', but *not* $Flies(Tweety)$, since the application of δ is "blocked":

- with
  - $Penguin(Tweety) \in E'$
  - and $\forall x(Penguin(x) \rightarrow \neg Flies(x)) \in E'$,

  it follows that $\neg Flies(Tweety) \in E'$ (since E' is closed under classical logic).

  ➥ $Flies(Tweety)$ can no longer be consistently assumed!

➥ It should hold that $E' = Cn(W')$.

☞ Note the nonmonotonic behaviour: although $W \subset W'$, we have $E \not\subseteq E'$ (since $Flies(Tweety) \in E$ but $Flies(Tweety) \notin E'$)!

Example: Nixon Diamond

- Consider the following default theory $T = (W, \Delta)$:

$$\begin{aligned} W &= \{Republican(Nixon), Quaker(Nixon)\}; \\ \Delta &= \left\{ \delta_1 = \frac{Republican(x) \ : \ \neg Pacifist(x)}{\neg Pacifist(x)}, \right. \\ &\qquad \left. \delta_2 = \frac{Quaker(x) \ : \ Pacifist(x)}{Pacifist(x)} \right\}, \end{aligned}$$

representing the following commonsense knowledge:

  - Nixon is both a Republican and a Quaker.
  - Republicans are normally not pacifists.
  - Quakers are normally pacifists.

- Defaults δ_1 and δ_2 are mutually conflicting! (The application of δ_1 blocks δ_2 and vice versa.)

Example: Nixon Diamond (ctd.)

- Thus, there are two alternatives for extensions:

$$E_1 = Cn(\{Republican(Nixon), Quaker(Nixon), \neg Pacifist(Nixon)\});$$
$$E_2 = Cn(\{Republican(Nixon), Quaker(Nixon), Pacifist(Nixon)\}).$$

☞ Formal definition of an extension follows next!