Defeasible Logic

1. G. Antoniou, D. Billington, G. Governatori and M. Maher. Representation Results for Defeasible Logic. *ACM Transactions on Computational Logic*, 2, 255-287, 2001.

2. M. Maher. Propositional Defeasible Logic has Linear Complexity. *Theory and Practice of Logic Programming*, 1 (6) 691-711, 2001.

3. D. Nute. Defeasible Logic. In D. M Gabbay, C. J. Hogger and J. A. Robinson (eds.): *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, Oxford University Press, 1994, 353-395.

Description of Defeasible Logic

DL – a non-monotonic formalism based on rules and a priority relation on them.

A theory in this logic consists of five components: *facts, strict rules, defeasible rules, defeaters and a superiority relation.*

Facts – classical logic statements describing indisputable facts, for example "Marco is Italian"

$$Italian(Marco) \tag{1}$$

or "Marco is a communist"

$$Communist(Marco). \tag{2}$$

Strict rules – rules in a classical sense, i.e. if premises of a rule hold then so do the conclusion. An example:

$$Italian(x) \rightarrow European(x). \tag{3}$$

Defeasible rules are understood in the following way: if premises of a rule hold and the contrary of the conclusion of the rule is unknown, then the conclusion holds.

An example: "Typically, Italians are catholics" represented as

$$Italian(x) \Rightarrow Catholic(x). \tag{4}$$

The rule (4) together with (1) allows us to conclude that Marco is a catholic.

Consider now the rule

$$Communist(x) \Rightarrow \neg Catholic(x). \tag{5}$$

Given (5) and (2) we can conclude that Marco is not a catholic. However, if the theory under consideration consists of (1), (2), (4) and (5), we cannot determine whether Marco is a catholic or not.

Defeaters – rules that are not used to draw any conclusions, but to prevent conclusions of some defeasible rules. An example:

$$Communist(x) \rightsquigarrow \neg Catholic(x). \tag{6}$$

The intuitive meaning: "for any object x, if x is a communist, then the conclusion that x is a catholic should be blocked."

From (2) and (6) we cannot conclude that Marco is not a catholic.

However, in the context with (1) and (4), the rule (6) prevent the conclusion that $Marco$ is a catholic.

The superiority relation among rules is used to make one rule stronger than another.

Assumptions:

1. the relation is cycles free
2. is defined *locally*, i.e. between rules with complementary heads.

For example, consider rules (1), (2), (4) and (5) and the superiority relation (5) $>$ (4). Now, we may conclude that Marco is not a catholic, because the conclusion of (4) is overriden by the conclusion of (5).

Formal definition

Language

A fixed subset of first-order language containing a finite set of constants and a finite set of relations.

Both facts and rules may contain free variables. Such facts (rules) are interpreted as schemata of facts (rules), i.e. as the set of all instances over the set of constants of the considered language.

A *defeasible theory* D – a triple $\langle F, R, > \rangle$, where

- F is a finite set of literals (called *facts*),
- R is a finite set of rules,
- $>$ is an acyclic superiority relation on R.

Rules are of the form:

$$r : A(r) \hookrightarrow C(r),$$

where

- r is a unique *label* of the rule,
- $A(r)$ – its *antecedent* (body) – a finite list of literals,
- $C(r)$ – its *consequent* (head) – a single literal,

- $\hookrightarrow$ depends on a kind of a rule:

$$\hookrightarrow = \begin{cases} \rightarrow & \text{if it is a strict rule} \\ \Rightarrow & \text{if it is a defeasible rule} \\ \rightsquigarrow & \text{if it is a defeater} \end{cases}$$

A superiority relation $>$ is a binary relation on R. $r_1 > r_2$ states that r_1 overrules r_2 if both rules are applicable. The relation $>$ is assumed to be acyclic and is defined between rules with complementary heads.

Proof theory

DL is defined by a proof theory.

R_s, R_d, R_{dft} – the subsets of strict rules, defeasible rules and defeaters of R, respectively.

$R[q]$ – a set of all rules from R with q in their heads.

A *conclusion* of a theory D is a tagged literal in one of the forms:

$+\mathbf{\Delta q}$ – means that q is definitely provable in D.

$-\mathbf{\Delta q}$ – means that we have proved that q is not definitely provable in D.

$+\delta\mathbf{q}$ – means that q is defeasibly provable in D.

$-\delta\mathbf{q}$ – means that we have proved that q is not defeasibly provable in D.

Provability in DL

The notion of provability is based on the concept of a *derivation* in $D = \langle F, R, > \rangle$ – a finite sequence $P = (P(1), \ldots, P(n))$ of tagged literals constructed by inference rules of four kinds.

$P(1..i)$ – the initial part of the sequence P of length i.

$+\Delta$: We may append $P(i+1) = +\Delta q$ if either
(1) $q \in F$ or
(2) $\exists r \in R_s[q]\ \forall a \in A(r) : +\Delta a \in P(1..i)$.

$-\Delta$: We may append $P(i+1) = -\Delta q$ if
(1) $q \notin F$ and
(2) $\forall r \in R_s[q]\ \exists a \in A(r) : -\Delta a \in P(1..i)$.

$+\delta$: We may append $P(i+1) = +\delta q$ if either

(1) $+\Delta q \in P(1..i)$ or

(2) (2.1) $\exists r \in R_{sd}[q]\ \forall a \in A(r) : +\delta a \in P(1..i)$ and

(2.2) $-\Delta\neg q \in P(1..i)$ and

(2.3) $\forall s \in R[\neg q]$ either

(2.3.1) $\exists a \in A(s) : -\delta a \in P(1..i)$ or

(2.3.2) $\exists t \in R_{sd}[q]$ such that $\forall a \in A(t) : +\delta a \in P(1..i)$ and $t > s$.

$-\delta$: We may append $P(i+1) = -\delta q$

(1) $-\Delta q \in P(1..i)$ and

(2) (2.1) $\forall r \in R_{sd}[q] \exists a \in A(r) : -\delta a \in P(1..i)$ or

(2.2) $+\Delta\neg q \in P(1..i)$ or

(2.3) $\exists s \in R[\neg q]$ such that

(2.3.1) $\forall a \in A(s) : +\delta a \in P(1..i)$ and

(2.3.2) $\forall t \in R_{sd}[q]$ either $\exists a \in A(t) : -\delta a \in P(1..i)$ or $t \not> s$.

An analisis of the proof theory

We may define four sets of literals wrt the types of conclusions of a theory D.

$$
\begin{aligned}
+\Delta &= \{p \mid D \vdash +\Delta p\} \\
-\Delta &= \{p \mid D \vdash -\Delta p\} \\
+\delta &= \{p \mid D \vdash +\delta p\} \\
-\delta &= \{p \mid D \vdash -\delta p\}
\end{aligned}
$$

As straightforward consequences of the proof rules, we have the following relations among the sets.

$$
\begin{aligned}
+\Delta &\subseteq +\delta \qquad & +\Delta \cap -\Delta &= \emptyset \\
-\delta &\subseteq -\Delta \qquad & +\delta \cap -\delta &= \emptyset
\end{aligned}
$$

Thus, there are six different possible outcomes.

A: $\not\vdash -\Delta p$ and $\not\vdash +\delta p$
$p \mapsto p$

B: $\vdash +\delta p$ and $\not\vdash +\Delta p$ and $\not\vdash -\Delta p$
$\Rightarrow p;\ p \mapsto p$

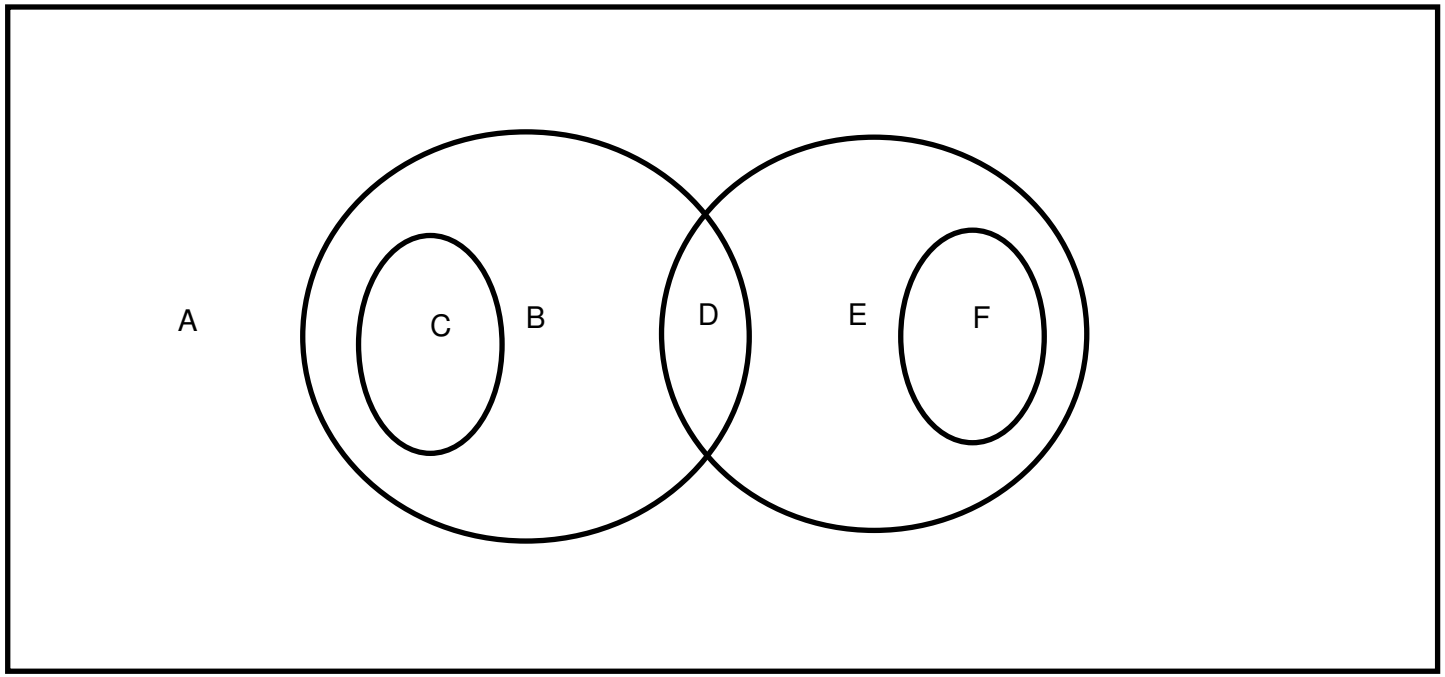
C: $\vdash +\Delta p$ (and $\vdash +\delta p$)
$\mapsto p$

D: $\vdash +\delta p$ and $\vdash -\Delta p$
$\Rightarrow p$

E: $\vdash -\Delta p$ and $\not\vdash= \delta p$ and $\not\vdash -\delta p$
$p \Rightarrow p$

F: $\vdash -\delta p$ (and $\vdash -\Delta p$)
$\emptyset$

These outcomes may be represented by the following diagram.



Proposition 1
Consider a defeasible theory D.

1. If $\nvdash -\Delta\neg p$ and $\nvdash +\Delta p$ then $\nvdash +\delta p$
2. If $\vdash +\Delta\neg p$ and $\vdash -\Delta p$ then $\vdash -\delta p$
3. If $\vdash +\delta\neg p$ and $\vdash -\Delta p$ and $\nvdash -\delta p$ then D is cyclic.∎

Theorem 1
The consequences of a defeasible theory D can be computed in $O(N)$ time, where N is the number of symbols in D.

Examples

Example 1

Consider a theory $\langle F, R, > \rangle$, where

$F = \{Italian(Marco), Communist(Marco)\}$

$R = \{r_1 : Italian(x) \Rightarrow Catholic(x), r_2 : Communist(x) \Rightarrow \neg Catholic(x)\}$

$> = \{r_2 > r_1\}$

Below is a derivation of $+\delta\neg Catholic(Marco)$.

$P = (+\Delta Italian(Marco), +\delta Italian(Marco), +\Delta Communist(Marco),$

$+\delta Communist(Marco), -\Delta Catholic(Marco), +\delta\neg Catholic(Marco)).$

Example 2

Consider the following scenario.

John is 17-year-old.

Typically, 17-year-olds are unmarried.

Typically, adults are married.

Typically, adults are employed.

Typically, employees are adult.

Given the above facts, we intuitively feel that "John is unmarried" should be provable. However, this conclusion cannot be derived in DL.

A DL theory corresponding to the scenario. (Here j, $Y17$, A, E and M stand for $John$, $17-year-old$, $Adult$, $Employed$ and $Married$, respectively.)

$Y17(j)$

$A(x) \Rightarrow E(x)$

$E(x) \Rightarrow A(x)$

$A(x) \Rightarrow M(x)$

$Y17(x) \Rightarrow \neg M(x)$.

The conclusion $+\delta\neg M(j)$ cannot be derived. ▪