NICTA

The Hardness of Revising Defeasible Preferences

Guido Governatori, Francesco Olivieri,
Simone Scannapieco, Matteo Cristani

RuleML 2014 Prague, August 20th 2014

Australian Government
Department of Broadband, Communications and the Digital Economy
Australian Research Council

NICTA Funding and Supporting Members and Partners
Australian National University · UNSW The University of New South Wales · NSW Government Trade & Investment · Victoria The Place To Be · The University of Melbourne · The University of Sydney · Queensland Government · Griffith University · QUT · The University Of Queensland Australia

Scope of the Paper

What preferences are?

- Means to choose a course of action to achieve a goal

We focus on a computational study of preference revision

- Fundamental in non-monotonic reasoning with partial/conflicting information $\rightarrow$ knowledge subject to continuous changes
- Like legal domain

We study scenarios where preference modification is the only way to change the knowledge about the environment

- An average citizen has no power to change the Law, nor decide what norms are effective in a particular context. . .
- . . . but only argue what norms prevail over others
  - idea underlying legal principles like lex superior, posterior, specialis

Logical Framework: **DL**

NICTA

Three type of knowledge:

- **Facts** always true
- **Rules** as relationships among propositional atoms
  - **Strict** $cat(X) \rightarrow_{r_1} mammal(X)$
  - **Defeasible** $cat(X) \Rightarrow_{r_2} eatBirds(X)$
  - **Defeaters** $justFed(X) \rightsquigarrow_{r_3} \neg eatBirds(X)$
- **Preferences** over rules, e.g., $r_5 > r_4$

$$cat(X) \Rightarrow_{r_4} eatBirds(X)$$
$$domesticCat(X) \Rightarrow_{r_5} \neg eatBirds(X)$$

Defeasible theories as tuples $(F, R, >)$

Proving Conclusions in DL

1. Give an argument for the conclusion we want to prove
2. Consider all possible counterarguments to it
3. Rebut all counterarguments
   - Defeat the argument by a stronger one
   - Undercut the argument by showing that some of the premises do not hold

We use proof tags

- $+\partial p$ (p is defeasibly proved)
- $-\partial p$ (p is defeasibly refuted)

Where All Began. . .

When p is a **tautology** in this framework?

- Classical sense: p true in every interpretation
- In our case: $D \vdash +\partial p$ regardless to $>$

1. p is a fact. . . And?
2. (Defeasible) reasoning chains for p with no undercuts nor rebuttals. . . But there is more.
3. $$\begin{array}{lllll} \Rightarrow_{r_1} & l & \Rightarrow_{r_2} & \neg a & \\ & & \Rightarrow_{r_3} & a & \Rightarrow_{r_4} p \\ & & \Rightarrow_{r_5} & b & \Rightarrow_{r_6} p \\ \Rightarrow_{r_7} & \neg l & \Rightarrow_{r_8} & \neg b & \end{array}$$

How We Did It

NICTA

Preference Revision Problem

INSTANCE: A defeasible theory $D = (F, R, >)$, and a literal p.
QUESTION: Is it possible to change $>$ into $>'$ such that $D' = (F, R, >')$ and either

1. If $D \vdash +\partial p$, then $D' \vdash -\partial p$?
2. If $D \vdash -\partial p$, then $D' \vdash +\partial p$?

Reduction from 3-SAT exploiting the previous rule pattern

- Literals of the formula becomes literals of a defeasible theory
- Each clause is transformed into a set of defeasible rules involving a fixed literal p

To Be More Specific. . .

Given a 3-SAT formula $\Gamma = \bigwedge_{i=1}^{n} C_i$ such that $C_i = \bigvee_{j=1}^{3} a_j^i$, we define the Γ**-transformation** as the operation that maps Γ into

$$\begin{aligned} R^{\Gamma} = \{ & r_{ij}^{a} : \Rightarrow a_j^i \\ & r_{ij} : a_j^i \Rightarrow c_i \\ & r_{\sim i} : \Rightarrow \sim c_i \\ & r_i : \sim c_i \Rightarrow p \}. \end{aligned}$$

We proved that:

- if p is tautological, then Γ is not satisfiable;
- if p is non-tautological, then Γ is satisfiable.

Conclusions

Our work comes full circle in computational study of defeasible revision

1. Revision of factual knowledge corresponds to update operation (Katsuno and Mendelzon, 1991)
2. Revision by changing rule set studied by Billington et al., 1999 (polynomial time)
3. Revision by changing preferences. . . Well, we just did it

Thank you for the attention!

Any question?