Better Automated Theorem Proving in Event-B

Matthias Schmalz

ETH Zurich

July 16, 2009

Outline

Rodin's Automated Theorem Provers (ATPs)

- Mono-Lemma Prover (Clearsy)
- Predicate Prover (Clearsy)
- New Predicate Prover (François Terrier, ETHZ)

Criticising an ATP is much easier than developing a better ATP!

Examples

First-Order Reasoning

$$c \in B$$
$$B = R$$
$$R \subseteq \mathrm{dom}(f)$$
$$\vdash_{\mathcal{L}}$$
$$c \in \mathrm{dom}(f)$$

New PP fails.

$$c \in B$$
$$B = R$$
$$R \subseteq \{y \mid \exists z \cdot y \mapsto z \notin (B \times B) \setminus f\}$$
$$\vdash_{\mathcal{L}}$$
$$c \in \mathrm{dom}(f)$$

New PP, PP, ML fail.

$\Rightarrow$ Provers are sensitive to the precise way of writing formulae.

Examples

Well-Definedness

~~$x \in \mathrm{dom}(f)$~~
~~$f \in A \nrightarrow A$~~
$\vdash_{\mathcal{L}}$
$f(x) \in \mathrm{ran}(f)$

New PP and PP fail.

$\Rightarrow$ Provers are unaware of Well-Definedness assumptions.

Examples

Arithmetic

Only ML succeeds:

$$\vdash_{\mathcal{L}} x * x \geq 0$$

$$x \in 1 \mathinner{..} 2 \vdash_{\mathcal{L}} (x - 1) * (x - 2) \leq 0$$

All ATPs fail:

$$\vdash_{\mathcal{L}} (x - 1) * (x - 3) \geq -1$$

$$x \in 1 \mathinner{..} 4 \vdash_{\mathcal{L}} (x - 1) * (x - 4) \leq 0$$

$\Rightarrow$ Arithmetic capabilities are hard to understand.

Examples

Consistency

$$\vdash_{\mathcal{L}} P = \text{TRUE} \lor Q = \text{TRUE} \Leftrightarrow P = \text{TRUE} \lor R = \text{TRUE}$$

Erroneously discharged by New PP

Problems

- ATPs do not discharge several “obvious” sequents.
- Strengths / limitations of ATPs are not well-understood.
- New PP is inconsistent.

My conjecture:

- Rodin’s ATPs do not scale as well as state of the art ATPs.

Outline

Possible Approaches

Improve New PP, PP or ML: ×

- PP and ML are closed source
- Risk of late failure

Develop new ATP: ×

- not enough time

Integrate an existing out-of-the box ATP: ✓

- There are ATP competitions (e.g. CASC).
- If this approach is doomed, I will realise early.

Short Term Plans

- Integrate an ATP from the CASC competition (e.g. the E prover).
- Develop a plug-in for evaluating Rodin's ATPs.
- Tune the integration based on numerous case studies.
- Gain a better understanding of the strengths and limitations of the CASC prover.

Integrate External ATP

Challenge: Translation to Plain Predicate Calculus

Original:

$G \subseteq C$
$H \subseteq C$
$f \in \mathbb{P}(C) \rightarrow D$
$\vdash_{\mathcal{L}}$
$G \cup H \in \mathrm{dom}(f)$

Translated:

$x \in G \Rightarrow x \in C$
$\ldots$
$\ldots$
$\vdash$
$\exists y \cdot \exists L \cdot (\forall x \cdot x \in L \Leftrightarrow x \in G \vee x \in H) \wedge$
$L \mapsto y \in f$

New PP and PP fail.

Translation introduces a non-trivial quantification over a set.
Translation makes the sequent unprovable.

What is a Good Translation?

A good translation . . .

- is sound,
- can be computed efficiently,
- some extent preserves provability.
- Limitations should be clear.

Long Term Plans

Possible directions:

- Improve arithmetic reasoning.
- Improve mechanisms for removing irrelevant hypotheses.
- Detect hypotheses used in an (automated) proof.
- Give useful feedback on unprovable sequents.

Expected (Practical) Contributions

Make Rodin's ATPs more usable from the user's perspective:

- Integrate more powerful ATPs into Rodin.
- Make the integration itself more effective.
- Make Rodin's ATPs succeed on more sequents that the user perceives as obvious.
- Make the strengths and limitations of the ATPs transparent to the user.
- Maybe: compute helpful feedback on unprovable sequents.