Term-Indexing

Alexander Maringele
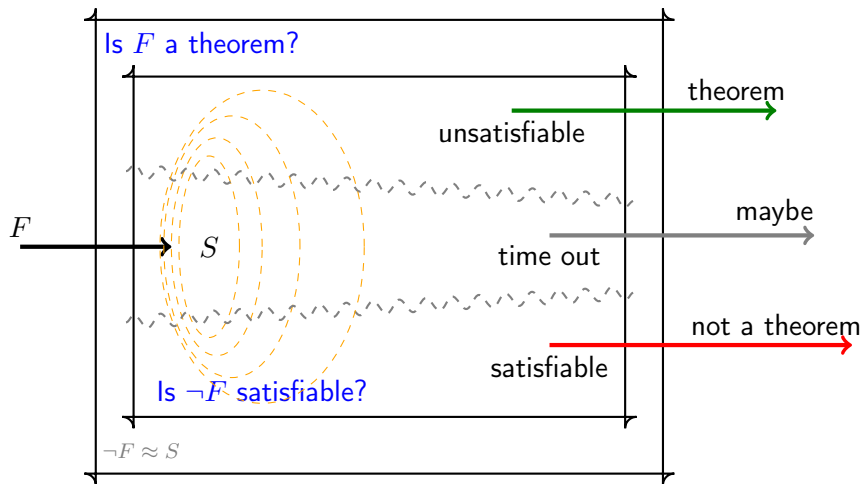
January 27th, 2016

References

R. Sekar, I. V. Ramakrishnan, and Andrei Voronkov, *Term indexing*, Handbook of Automated Reasoning (Alan Robinson and Andrei Voronkov, eds.), Elsevier Science Publishers B. V., Amsterdam, The Netherlands, 2001, pp. 1853–1964.

Outline

Refutation

Notation

> **Clausal form**
>
> $$\{ \, \mathsf{P}(\mathsf{f}(x)) \vee \mathsf{f}(x) \not\approx \mathsf{a}, \; \mathsf{g}(x,y) \approx \mathsf{a} \vee \neg\mathsf{Q}(x,y), \; \mathcal{C}_3 \, \}$$
> $$\equiv$$
> $$\forall x \, (\mathsf{P}(\mathsf{f}(x)) \vee \mathsf{f}(x) \not\approx \mathsf{a})$$
> $$\wedge$$
> $$\forall xy \, (\mathsf{g}(x,y) \approx \mathsf{a} \vee \neg\mathsf{Q}(x,y))$$
> $$\wedge$$
> $$\forall \, \mathcal{V}\mathsf{ar}(\mathcal{C}_3) \, (\mathcal{C}_3)$$

Goal

A sound and refutation complete calculus.

Resolution (without equality)

Resolve and factor all clauses and literals in an unsatisfiable set

$$\frac{A \vee \mathcal{C} \quad \neg B \vee \mathcal{D}}{(\mathcal{C} \vee \mathcal{D})\sigma} \; (\sigma) \text{ resolution} \qquad \frac{A \vee B \vee \mathcal{C}}{(A \vee \mathcal{C})\sigma} \; (\sigma) \text{ factoring}$$

$$\sigma = \mathrm{mgu}(A, B)$$

and the empty clause will be derived eventually.

Observation

Usually the set grows too fast to obtain a result.

Goal

A sound, refutation complete, and *effective* calculus.

1. *Reduce* search space
   - Ordered Resolution, Strategies, . . .
   - . . . with selection functions for clauses and literals
2. *Reduce* redundancy
   - e.g. discard clauses that are subsumed by other clauses
   - . . . depending on the calculus

Example (forward subsumption)

$$S = \{ \overset{1:}{\mathsf{P}}(x,y), \overset{2:}{\neg}\mathsf{P}(\mathsf{a},z) \} \cup \{ \overset{3:}{\mathsf{P}}(\mathsf{a},z') \} \qquad t_1 \text{ subsumes } t_3$$

$$\frac{\mathsf{P}(x,y) \quad \neg\mathsf{P}(\mathsf{a},z)}{\square} \ \{ x \mapsto \mathsf{a}, y \mapsto z \} \qquad \text{Resolution}$$

$$S\perp = \{ \mathsf{P}(\perp,\perp), \neg\mathsf{P}(\mathsf{a},\perp), \mathsf{P}(\mathsf{a},\perp) \} \qquad \text{InstGen / SMT}$$

Goal

A sound, refutation complete, and effective calculus.

3 Quickly find
  - *variants*   variant removal
  - *instances*   backward subsumption
  - *generalizations*   forward subsumption
  - *unifiable terms*   resolution, demodulation

  of a query term in a given set of terms.

Observation

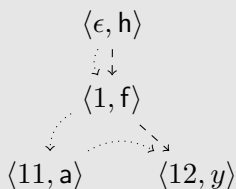Deduction rate drops quickly with sequential search.

Term Indexing

Data structures and algorithms for fast retrieval of matching terms.

Position Strings

Definition

$$\mathcal{P}\mathsf{os}^{\Sigma}(t) = \begin{cases} \{\langle \epsilon, x \rangle\} & \text{if } t = x \in \mathcal{V} \\ \{\langle \epsilon, f \rangle\} \cup \{\langle ip, s \rangle \mid \langle p, s \rangle \in \mathcal{P}\mathsf{os}^{\Sigma}(t_i)\} & \text{if } t = f(t_1, \ldots, t_n) \end{cases}$$

Term traversals

$\langle \epsilon, \mathsf{h} \rangle$

$\langle 1, \mathsf{f} \rangle$

$\langle 11, \mathsf{a} \rangle \qquad \langle 12, y \rangle$

$\mathcal{P}\mathsf{os}^{\Sigma}(\mathsf{h}(\mathsf{f}(\mathsf{a}, y))) = \{\langle \epsilon, \mathsf{h} \rangle, \langle 1, \mathsf{f} \rangle, \langle 11, \mathsf{a} \rangle, \langle 12, y \rangle\}$

$\langle \epsilon, \mathsf{h} \rangle \langle 1, \mathsf{f} \rangle \langle 12, y \rangle$ path from root to leaf

$\langle \epsilon, \mathsf{h} \rangle \langle 1, \mathsf{f} \rangle \langle 11, \mathsf{a} \rangle \langle 12, y \rangle$ pre-order traversal

Variables

Variants of terms generate the same position strings

- if variable names are ignored $\qquad$ $f(y, z) \Rightarrow \langle \epsilon, f \rangle \langle 1, * \rangle \langle 2, * \rangle$
- or normalized $\qquad f(y, z) \Rightarrow \langle \epsilon, f \rangle \langle 1, x_1 \rangle \langle 2, x_2 \rangle$
  $\qquad f(y, y) \Rightarrow \langle \epsilon, f \rangle \langle 1, x_1 \rangle \langle 2, x_1 \rangle$

In the first case even non-variants of terms generate the same strings.

Notation

We abbreviate

- path strings $\langle \epsilon, h \rangle \langle 1, f \rangle \langle 12, * \rangle$ $\qquad$ h.1.f.2.$*$
- and traversal strings $\langle \epsilon, h \rangle \langle 1, f \rangle \langle 11, * \rangle \langle 12, * \rangle$ $\qquad$ h.f.a.$*$
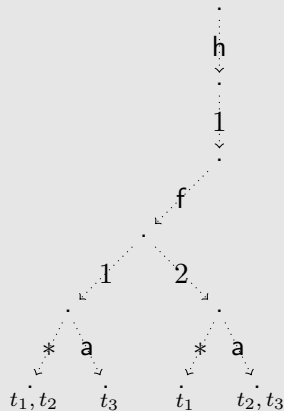 when traversal order and arities of symbols are fixed.

Path Indexing

Build

$$^{t_1:}\mathsf{h}(\mathsf{f}(x,y)),\,^{t_2:}\mathsf{h}(\mathsf{f}(x,\mathsf{a})),\,^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{a},\mathsf{a}))$$

$$t_1 \Rightarrow \{\mathsf{h.1.f.1.*}, \mathsf{h.1.f.2.*}\}$$
$$t_2 \Rightarrow \{\mathsf{h.1.f.1.*}, \mathsf{h.1.f.2.a}\}$$
$$t_3 \Rightarrow \{\mathsf{h.1.f.1.a}, \mathsf{h.1.f.2a}\}$$

Retrieve

$^{t_1:}\mathsf{h}(\mathsf{f}(x,y)),\,^{t_2:}\mathsf{h}(\mathsf{f}(x,\mathsf{a})),\,^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{a},\mathsf{a}))$

$$\mathsf{h}(\mathsf{f}(z,\mathsf{b}))) \Rightarrow \{\mathsf{h}.\mathsf{f}.*,\mathsf{h}.\mathsf{f}.\mathsf{b}\}$$

$u : \mathsf{h}(\mathsf{f}(z,\mathsf{b})) \mapsto \{t_1,t_2,t_3\} \cap \{t_1\}$

$i : \mathsf{h}(\mathsf{f}(z,\mathsf{b})) \mapsto \{t_1,t_2,t_3\} \cap \{\}$

$g : \mathsf{h}(\mathsf{f}(z,\mathsf{b})) \mapsto \{t_1,t_2\} \cap \{t_1\}$

$v : \mathsf{h}(\mathsf{f}(z,\mathsf{b})) \mapsto \{t_1,t_2\} \cap \{\}$

$v : \mathsf{h}(\mathsf{f}(z,z)) \mapsto \{t_1,t_2\} \cap \{t_1\}$
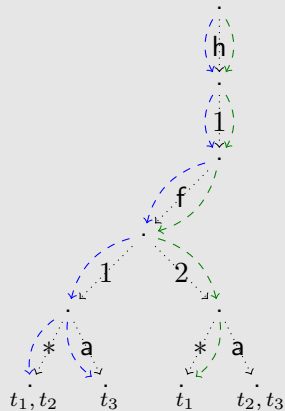
Unit Superposition Inference Rules

$$\frac{s \approx t \quad L[s']}{(L[t]) \cdot \sigma} \quad \text{unit paramodulation}$$

where $\sigma = \mathrm{mgu}(s, s')$, $s' \notin \mathcal{V}$, $t\sigma \not\succ s\sigma$

$$\frac{s \approx t \quad u[s'] \not\approx v}{(u[t] \not\approx v) \cdot \sigma} \quad \text{unit superposition} \qquad \frac{s \approx t \quad u[s'] \approx v}{(u[t] \approx v) \cdot \sigma}$$
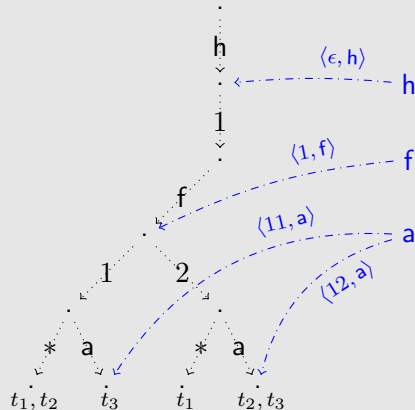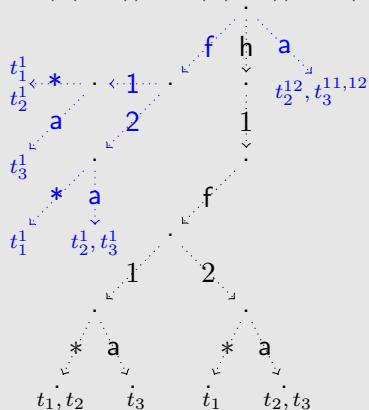
where $\sigma = \mathrm{mgu}(s, s')$, $s' \notin \mathcal{V}$, $t\sigma \not\succ s\sigma$, $v\sigma \not\succ u[s']\sigma$

$$\frac{s \not\approx t}{\Box} \quad \text{unit equality resolution} \qquad\qquad \frac{A \quad \neg B}{\Box} \quad \text{unit resolution}$$

where s and t (A and B respectively) are unifiable

$^{t_4:}f(x, a) \approx x$

$^{t_1:}h(f(x, y)), ^{t_2:}h(f(x, a)), ^{t_3:}h(f(a, a))$

Discrimination Trees

Insert

${}^{t_1:}\mathsf{h}(\mathsf{f}(x,y)), {}^{t_2:}\mathsf{h}(\mathsf{f}(x,\mathsf{h}(\mathsf{a}))), {}^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{h}(\mathsf{a}),\mathsf{a}))$

$$t_1 \Rightarrow \mathsf{h.f.*.*}$$
$$t_2 \Rightarrow \mathsf{h.f.*.h.a}$$
$$t_3 \Rightarrow \mathsf{h.f.h.a.a}$$
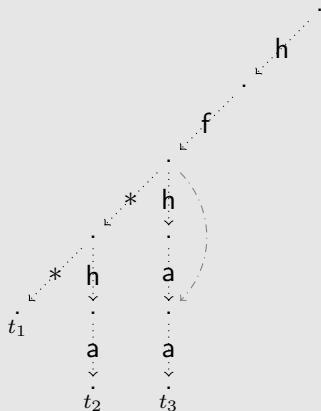
Retrieve

$${}^{t_1:}\mathsf{h}(\mathsf{f}(x,y)),{}^{t_2:}\mathsf{h}(\mathsf{f}(x,\mathsf{h}(\mathsf{a}))),{}^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{h}(\mathsf{a}),\mathsf{a}))$$

$$\mathsf{h}(\mathsf{f}(x',\mathsf{a})) \Rightarrow \mathsf{h}.\mathsf{f}.*.\mathsf{a}$$

$$u : \mathsf{h}(\mathsf{f}(x',\mathsf{a})) \mapsto \{t_1, t_3\}$$
$$i : \mathsf{h}(\mathsf{f}(x',\mathsf{a})) \mapsto \{t_3\}$$
$$g : \mathsf{h}(\mathsf{f}(x',\mathsf{a})) \mapsto \{t_1\}$$
$$v : \mathsf{h}(\mathsf{f}(x',\mathsf{a})) \mapsto \{\ \}$$

Subterms
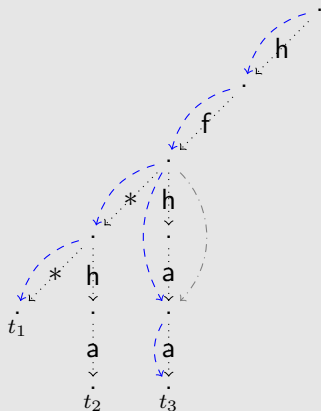
$^{t_1:}\mathsf{h}(\mathsf{f}(x,y)),\,^{t_2:}\mathsf{h}(\mathsf{f}(x,\mathsf{h}(\mathsf{a}))),\,^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{h}(\mathsf{a}),\mathsf{a}))$
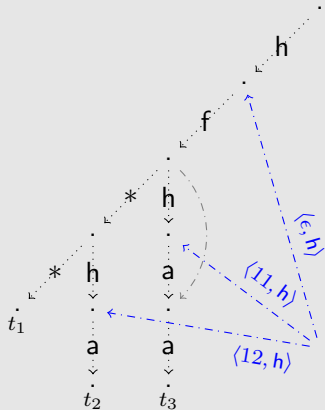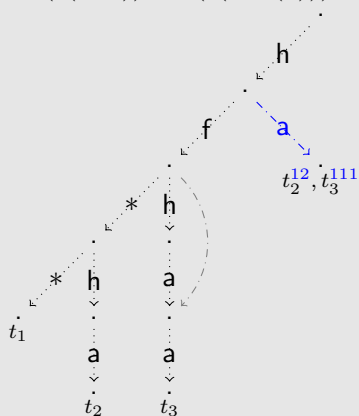
Substitution Trees

Build

$${}^{t_1:}h(f(x,y)), {}^{t_2:}h(f(x,h(a))), {}^{t_3:}h(f(h(a),a)), {}^{t_4:}h(f(a,a))$$

$$\downarrow$$

$$*_0 \mapsto h(*_1)$$

$$\downarrow$$

$$*_1 \mapsto f(*_2, *_3)$$

$$*_2 \mapsto x \qquad\qquad *_3 \mapsto a$$

$$*_3 \mapsto y \qquad *_3 \mapsto h(a) \qquad *_2 \mapsto h(a) \qquad *_2 \mapsto a$$
$$t_1 \qquad\qquad t_2 \qquad\qquad t_3 \qquad\qquad t_4$$

```
TPTP/Problems/HWV/HWV134-1.p
2 332 428 formulae, 6 570 884 literals
```

checking afterwards	1000 new literals (ℓ_1, ℓ_2)	$A, \neg B$	sequential search	path index	speed up
1 000	500 000	761	726ms	70ms	10
2 000	1 500 000	812	2s	69ms	29
4 000	3 500 000	723	4s	75ms	53
8 000	7 500 000	433	9s	125ms	72
16 000	15 500 000	742	21s	221ms	95
32 000	31 500 000	592	40s	489ms	82
64 000	63 500 000	1 167	*80s*	697ms	115
128 000	127 500 000	1 479	*160s*	13s	12
256 000	255 500 000	1 097	*320s*	440s	1
512 000	511 500 000	1 440	*640s*	348s	2
1 024 000	1023 500 000	1 534	*1280s*	330s	4