Duplicate Check

Nao Hirokawa

JAIST

February 20, 2024

About This Talk

1. SMT solving for $\mathcal{R} \approx \mathcal{S}$ (based on Thiemann's approach)
2. fingerprints (variant of Felgenhauer's approach)
3. experiments
4. future work

Example (partition)

consider collection of TRSs $\mathfrak{R} = \{\mathcal{R}_1, \ldots, \mathcal{R}_6\}$:

$$\mathcal{R}_1 = \left\{ \begin{array}{c} +(0, x) \to x \\ +(\mathsf{s}(x), y) \to \mathsf{s}(+(x, y)) \\ \mathsf{p}(\mathsf{s}(x)) \to x \end{array} \right\} \qquad \mathcal{R}_3 = \{\mathsf{a} \to \mathsf{b}\} \qquad \mathcal{R}_4 = \{\mathsf{d} \to \mathsf{c}\}$$

$$\mathcal{R}_2 = \left\{ \begin{array}{c} \mathsf{plus}(\mathsf{zero}, n) \to n \\ \mathsf{plus}(\mathsf{succ}(m), n) \to \mathsf{succ}(\mathsf{plus}(m, n)) \\ \mathtt{pred}(\mathsf{succ}(n)) \to n \end{array} \right\} \qquad \mathcal{R}_5 = \{\mathsf{c} \to \mathsf{d}\} \qquad \mathcal{R}_6 = \{\mathsf{f}(\mathsf{f}(x)) \to x\}$$

duplicate checker is expected to compute partition by $\approx$:

$$\mathfrak{R}/{\approx} \;=\; \{\{\mathcal{R}_1, \mathcal{R}_2\}, \{\mathcal{R}_3, \mathcal{R}_4, \mathcal{R}_5\}, \{\mathcal{R}_6\}\}$$

Definition (definition of duplicates)

- $\mathcal{R} \doteq \mathcal{S}$ if $\mathcal{R}$ and $\mathcal{S}$ are same modulo variable renaming
- $\mathcal{R} \approx \mathcal{S}$ if $\mathcal{R}^\theta \doteq \mathcal{S}$ for some renaming θ for function symbols

Example

$$\left\{ \begin{array}{c} \mathsf{plus}(\mathsf{zero}, n) \to n \\ \mathsf{plus}(\mathsf{succ}(m), n) \to \mathsf{succ}(\mathsf{plus}(m, n)) \\ \mathtt{pred}(\mathsf{succ}(n)) \to n \end{array} \right\} \approx \left\{ \begin{array}{c} +(0, x) \to x \\ +(\mathsf{s}(x), y) \to \mathsf{s}(+(x, y)) \\ \mathsf{p}(\mathsf{s}(x)) \to x \end{array} \right\}$$

because of renaming θ: $\left\{ \begin{array}{c} \mathsf{plus}^\theta = + \\ \mathsf{succ}^\theta = \mathsf{s} \\ \mathsf{pred}^\theta = \mathsf{p} \end{array} \right\}$

Q. how to find it? A. reducing problem to equality constraint

Encoding Equivalence of Rules

Example

$$\begin{array}{rl} & \mathsf{plus}^\theta(\mathsf{succ}^\theta(x_1),x_2) \to \mathsf{succ}^\theta(\mathsf{plus}^\theta(x_1,x_2)) \\ = & +(\mathsf{s}(x_1),x_2) \to \mathsf{s}(+(x_1,x_2)) \end{array} \iff \bigwedge \begin{Bmatrix} \mathsf{plus}^\theta = + \\ \mathsf{succ}^\theta = \mathsf{s} \\ \mathsf{succ}^\theta = \mathsf{s} \\ \mathsf{plus}^\theta = + \end{Bmatrix}$$

$$\begin{array}{rl} & \mathsf{plus}^\theta(\mathsf{zero}^\theta, x_1) \to x_1 \\ = & +(\mathsf{s}(x_1),x_2) \to \mathsf{s}(+(x_1,x_2)) \end{array} \iff \bigwedge \begin{Bmatrix} \mathsf{plus}^\theta = + \\ \bot \end{Bmatrix}$$

$$\begin{array}{rl} & \mathsf{f}^\theta(x_1) \to x_2 \ \mathit{if}\ \mathsf{g}^\theta(x_1) \to x_2 \\ = & \mathsf{g}(x_1) \to x_2 \ \mathit{if}\ \mathsf{f}(x_1) \to x_2 \end{array} \iff \bigwedge \begin{Bmatrix} \mathsf{f}^\theta = \mathsf{g} \\ \mathsf{g}^\theta = \mathsf{f} \end{Bmatrix}$$

Fact

$$\mathcal{R} \subseteq \mathcal{S} \iff \bigwedge_{\alpha \in \mathcal{R}} \bigvee_{\beta \in \mathcal{S}} \alpha = \beta \quad \text{and} \quad \mathcal{R}^\theta = \mathcal{S} \iff \mathcal{R}^\theta \subseteq \mathcal{S} \wedge \mathcal{R}^\theta \supseteq \mathcal{S}$$

Example

$$\begin{Bmatrix} \mathsf{plus}(\mathsf{zero}, x_1) \to x_1 \\ \mathsf{plus}(\mathsf{succ}(x_1),x_2) \to \mathsf{succ}(\mathsf{plus}(x_1,x_2)) \\ \mathtt{pred}(\mathsf{succ}(x_1)) \to x_1 \end{Bmatrix}^\theta \subseteq \begin{Bmatrix} +(0, x_1) \to x_1 \\ +(\mathsf{s}(x_1),x_2) \to \mathsf{s}(+(x_1,x_2)) \\ \mathsf{p}(\mathsf{s}(x_1)) \to x_1 \end{Bmatrix}$$

$$\iff \bigwedge \begin{Bmatrix} (\mathsf{plus}^\theta = + \wedge \mathsf{zero}^\theta = 0) \vee \bot \vee \bot \\ \bot \vee (\mathsf{plus}^\theta = + \wedge \mathsf{succ}^\theta = \mathsf{s}) \vee \bot \\ \bot \vee \bot \vee (\mathsf{pred}^\theta = \mathsf{p} \wedge \mathsf{succ}^\theta = \mathsf{s}) \end{Bmatrix}$$

Encoding Equivalence of Signatures

$$\begin{Bmatrix} (\mathsf{plus}, 2) \\ (\mathsf{succ}, 1) \\ (\mathsf{pred}, 1) \end{Bmatrix}^\theta \subseteq \begin{Bmatrix} (+, 2) \\ (\mathsf{s}, 1) \\ (\mathsf{p}, 1) \end{Bmatrix}$$

$$\iff \bigwedge \begin{Bmatrix} \mathsf{plus}^\theta = + & \vee & \bot & \vee & \bot \\ \bot & \vee & \mathsf{succ}^\theta = \mathsf{s} & \vee & \mathsf{succ}^\theta = \mathsf{p} \\ \bot & \vee & \mathsf{pred}^\theta = \mathsf{s} & \vee & \mathsf{pred}^\theta = \mathsf{p} \end{Bmatrix}$$

Note

- equivalence of $(\mathcal{F},\mathcal{R})$ and $(\mathcal{G},\mathcal{S})$ is decided by solving $\mathcal{F}^\theta = \mathcal{G} \wedge \mathcal{R}^\theta = \mathcal{S}$
- in same way, CS(C)TRSs and MSTRSs can be handled

Duplicate Check I

given $(\mathcal{F}_1,\mathcal{R}_1),\ldots,(\mathcal{F}_n,\mathcal{R}_n)$
union-find maintains equivalence on $\{1,\ldots,n\}$
for each (i,j) with $i < j$
  if find$(i) \neq$ find(j)
    if $\mathcal{F}_i^\theta = \mathcal{F}_j$ and $\mathcal{R}_i^\theta \doteq \mathcal{R}_j$ for some θ — takes 1ms
      union(i,j)

Estimation of Runtime

if $|\mathfrak{R}| = 1500$ then $\dfrac{1500 \times 1499}{2} \times 1$ ms ≈ 20 min

Idea

use fingerprint function φ: $\varphi(\mathcal{R}) \neq \varphi(\mathcal{S}) \implies \mathcal{R} \not\approx \mathcal{S}$

$\varphi(\mathcal{R})$ in Current Implementation

$$\mathcal{R} = \left\{ \begin{array}{c} \overset{0}{\mathsf{a}}(\overset{1}{\mathsf{b}}(\overset{2}{\mathsf{c}}(\overset{0}{x}, \overset{1}{y}))) \to \overset{1}{\mathsf{b}}(\overset{0}{\mathsf{a}}(\overset{2}{\mathsf{c}}(\overset{1}{y}, \overset{1}{y}))) \\ \overset{0}{\mathsf{a}}(\overset{0}{x}) \to \overset{1}{\mathsf{b}}(\overset{0}{x}) \\ \overset{0}{\mathsf{b}}(\overset{0}{x}) \to \overset{1}{\mathsf{a}}(\overset{0}{x}) \end{array} \right\}$$

$$\varphi(\mathcal{R}) = \mathsf{concat}(\mathsf{sort}(\begin{bmatrix} [0,1,2,0,1,\ 1,0,3,1,1], \\ [0,0,\ 1,0], \\ [0,0,\ 1,0] \end{bmatrix}))$$

$$= [0,0,1,0,\ \ 0,0,1,0,\ \ 0,1,2,0,1,1,0,3,1,1]$$

Theorem

$\varphi(\mathcal{R}) \neq \varphi(\mathcal{S}) \implies \mathcal{R} \not\approx \mathcal{S}$ if $\mathcal{R}$ and $\mathcal{S}$ are variant-free (why?)

Duplicate Check II

given $(\mathcal{F}_1, \mathcal{R}_1), \ldots, (\mathcal{F}_n, \mathcal{R}_n)$

union-find maintains equivalence on $\{1, \ldots, n\}$

for each i

  compute variant-free version $\mathcal{R}_i'$ of $\mathcal{R}_i$

  $\vec{v}_i = \varphi(\mathcal{R}_i')$

for each (i, j) with $i < j$

  if find$(i) \neq$ find(j)

    if $\vec{v}_i = \vec{v}_j$ — takes $<$ 0.01ms

      if $\mathcal{F}_i^\theta = \mathcal{F}_j$ and $\mathcal{R}_i^\theta \doteq \mathcal{R}_j$ for some θ — takes 1ms

        union(i, j)

Experimental Results

- Core i5-1340P @ 1.30GHz
- Z3 for finding renaming
- ARI contains 1218 (CS)(C)TRSs

DB	problems	duplicates	SMT calls	time (sec)
TRSs in COPS	577	13	16	**0.12**
ARI	1503	0	8	**0.37**

Remark

some TRS in COPS is not variant-free; what about ARI?

Future Work I: MSTRSs

$$\{\mathsf{a} : \mathsf{A},\ \mathsf{b} : \mathsf{B}\}^\theta \approx \{0 : \mathsf{C},\ 1 : \mathsf{D}\}$$

$$\iff \bigwedge \left\{ \begin{array}{l} (\mathsf{a}^\theta = 0 \ \wedge\ \mathsf{A}^\theta = \mathsf{C}) \ \vee\ (\mathsf{a}^\theta = 1 \ \wedge\ \mathsf{A}^\theta = \mathsf{D}) \\ (\mathsf{b}^\theta = 0 \ \wedge\ \mathsf{B}^\theta = \mathsf{C}) \ \vee\ (\mathsf{b}^\theta = 1 \ \wedge\ \mathsf{B}^\theta = \mathsf{D}) \end{array} \right\}$$

Future Work II: COM

indexes can be swapped

```
(format TRS :number 2)
(fun f 1)
(rule (f x) x     :index 1)
(rule (f (f x)) x :index 2)
```

is equivalent to

```
(format TRS :number 2)
(fun g 1)
(rule (g x) x     :index 2)
(rule (g (g x)) x :index 1)
```


Future Work III: INF

arguments can be swapped

```
(format TRS :problem infeasibility)
(fun a 0)
(fun b 0)
(infeasibility? (= a a) (= a b))
```

is equivalent to

```
(format TRS :problem infeasibility)
(fun a 0)
(fun b 0)
(infeasibility? (= a b) (= a a))
```

probably infeasibility problems are generated by tools

Future Work IV

- duplicate check for LCTRSs (no idea about notion of equivalence)
- how does termCOMP gourp manage problems in TPDB?
- is duplicate check in P? or NP-complete?
- how to verify correctness of duplicate checkers?

  in COPS, some CTRS was wrongly tagged

Thanks for Your Attention!