Shapes/Slotted E-graphs

Definition slotted e-graph

Similar to (regular) egraphs:

- function symbols f, g
- e-class ids a, b, c
- slots $s_1, s_2, \ldots$
- slotmap $m ::= [s_j \mapsto s_k, \ldots]$ bijection
- invocation $i ::= m * a$
- terms $t ::= f \mid f(t_1, \ldots, t_k) \mid s_j \mid \lambda s_j.t$
- e-nodes $n ::= f \mid f(i_1, \ldots, i_k) \mid s_j \mid \lambda s_j.i$
- e-classes $c ::= \{n_1, \ldots n_m\} :: \{s_{i_1}, \ldots, s_{i_k}\}$

We have a mapping $\text{Classes} : \text{Id} \rightarrow \text{Eclass}$ to interpret the invocations.

Additional definitions

slots(_)

We define a family of (overloaded) functions slots for e-class ids, invocations, terms and e-nodes to a set of slots $\{s_{i_1}, \ldots, s_{i_k}\}$.

slots(_) for E-Classes, Ids and Invocations

- $\text{slots}(\{n_1, \ldots n_m\} :: \{s_{i_1}, \ldots, s_{i_k}\}) := \{s_{i_1}, \ldots, s_{i_k}\}$
- $\text{slots}(a) := \text{slots}(Classes(a))$
- $\text{slots}(m * a) := m \circ \text{slots}(a) = \{m(s_j) \mid s_j \in \text{slots(a)}\}$

slots(_) for Terms and E-Nodes Let $x, x_1, \ldots$ be either terms t or invocations i.

- $\text{slots}(f) := \emptyset$
- $\text{slots}(f(x_1, \ldots, x_k)) := \text{slots}(x_1) \cup \ldots \cup \text{slots}(x_k)$
- $\text{slots}(s_j) := \{s_j\}$
- $\text{slots}(\lambda s_j.x) := \text{slots}(x) \setminus \{s_j\}$

slots(t) on terms corresponds to the set of free variables.

The Action m * _

We define a family of (overloaded) functions $m * \_$ for e-class ids, invocations, terms and e-nodes.

Generally, $m * x$ is only defined, if $\text{slots}(x) \subseteq \text{dom}(m)$.

- For any x we have $m * (m' * x) = (m' * m) * x$
- with $m * m' := m \circ m' = \{x \mapsto z \mid x \mapsto y \in m', y \mapsto z \in m\}$

m * __ for E-Classes, Ids and Invocations

- $m * \{n_1, \ldots, n_k\} :: \{s_{i_1}, \ldots, s_{i_l}\} = \{m * n_1, \ldots, m * n_k\} :: \{m * s_{i_1}, \ldots, m * s_{i_l}\}$
- $m * a$ is just $m * a$, there is no way to simplify it
- For Invocations $i = m * a$, we define $m' * i := (m' * m) * a$

Note that there is a difference in semantics between e-classes and e-class ids for this action. So it might be necessary to keep them apart.

m * __ for Terms and E-Nodes Let $x, x_1, \ldots$ be either terms t or invocations i.

- $m * f := f$
- $m * f(x_1, \ldots, x_k) := f(m * x_1, \ldots, m * x_k)$
- $m * s_j := m(s_j)$
- $m * (\lambda s_j.x) := \lambda s_j.(m * x)$, assuming s_j is neither in the domain nor codomain of m.

We follow the Barendregt convention: We assume that all bound slots are never colliding with anything else. And if they do, we just rename them.

(Note to future self: We also need the Barendregt convention for redundant slots)

We claim that this definition implies $\text{slots}(m * x) = m \circ \text{slots}(x)$ for all x.

Examples:

- $\text{slots}(\lambda s_1.f(s_1, s_2, s_3)) = \text{slots}(f(s_1, s_2, s_3)) \setminus \{s_1\} = \{s_2, s_3\}$
- $\lambda s_1.f(s_1, s_2, s_3) * (s_1 \mapsto s_2, s_2 \mapsto s_3, s_3 \mapsto s_1)$ does not typecheck
- $\lambda s_1.f(s_1, s_2, s_3) * (s_2 \mapsto s_3, s_3 \mapsto s_1) = \lambda s_1.f(s_1, s_2, s_3)$ needs freshness
- $[s_2 \mapsto s_3, s_{47} \mapsto s_2] * a = [s_47 \mapsto s_2, s_2 \mapsto s_3] * a; \text{slots}(a) = \{s_2, s_47\}$

Containment

We define an element relation $\in \subseteq \text{Enodes} \times \text{Invocations}$ as: - $n \in m * a$, iff $n \in m * Classes(a)$

Notes

- We need to be more precise about the slots of e-nodes and e-classes (union/intersection, etc)
- We allow the shortcut for ordered (instead of named) arguments, $[s_j, s_{j'}, s_{j''}, ...] * i := [s_k \mapsto s_j, s_{k'} \mapsto s_{j'}, s_{k''} \mapsto s_{j''}, ...] * i$, assuming $slots(i) = \{s_k, s_{k'}, s_{k''}, ...\}$ and $k < k' < k'' < ....$

Automorphism Group

The operator $*$ defines a left group action of the group $G \leq \text{Sym}(\text{slots}(x))$.

The automorphism group $\text{Aut}(c)$ of an e-class $c = \{n_1, \dots, n_m\} :: \{s_{i_1}, \dots, s_{i_k}\}$ is the largest subgroup $\text{Aut}(c) \leq \text{Sym}(s_{i_1}, \dots, s_{i_m})$, such that $m * c = c$ for all $m \in \text{Aut}(c)$.

Orbits, Canonical Elements

For an e-node n and a group of slotmaps $M \leq \text{slots}(n)$, the orbit $M * n$ is the set of all permutations of n according to the group M, i.e. $M * n = \{m * n \mid m \in M\}$. Given a term ordering (we assume lexicographical) $<$, we define a canonical element of the orbit $M * n := \min_{m \in M} m * n$ to be the minimal representative of the orbit.

Weak Shapes

We define the weak shape of an e-node n as follows:

- $\text{weak\_shape}(n) := \min_{\theta \in \text{Sym}(S)} \{\theta * n\}$, where $S = \{s_j \mid j \in \mathbb{N}\}$ is the set of all slots.

Example

- Consider the e-class $c = s0 + s1, s1 + s0 :: s0, s1$, then the two e-nodes $f([s2, s3] * c, [s2, s3] * c)$ and $f([s2, s3] * c, [s3, s2] * c)$ should have the same hash because $[s2, s3] * c = [s3, s2] * c$. However, they don't have the same weak shape, because we don't compute the (weak shapes) of the invocations $[s2, s3] * c$.

Strong shape

Definition $\text{strong\_shape}(f(m_1 * c_1, \dots, m_k * c_k)) :=$

$$\min\{\text{weak\_shape}(f(m_1 * m_1' * c_1, \dots, m_k * m_k' * c_k)) \mid m_i' \in \text{Aut}(c_i)\},$$

or equivalently

$$\min\{\theta * f(m_1 * m_1' * c_1, \dots, m_k * m_k' * c_k) \mid m_i' \in \text{Aut}(c_i),\ \theta \in \text{Sym}(S)\}$$

- This typechecks because $\text{slots}(m * c) = \text{slots}(c)$ for all $m \in \text{Aut}(c)$
- enodes must be hashable
- hash must be invariant of renamings
- weak shape: canonical naming $s_1, s_2, \dots$
- egraph idea: congruence, i.e. if $a = b \Rightarrow f(a) = f(b)$ (in memory). This does not work in weak shapes:

- example: $\text{Classes}(a) = \{(fs_10s_11), (fs_11s_10)\} :: \{s_10, s_11\}$ encodes: $(+[s_10 \to s_1, s_11 \to s_2] * a[s_10 \to s_2, s_11 \to s_1] * a)$ and $(+[s_10 \to s_1, s_11 \to s_2] * a[s_10 \to s_1, s_11 \to s_2] * a)$ concrete terms correspond to $ (+ (f x y) (f y x))$ and $(+(fxy)(fxy))$
- strong shape: lex-min of all equivalent weak shapes.
- Conjecture: this is the double coset contstructive orbit problem

Divide and conquer doesn't work because:

$$\min\{\theta * m_2 * m_2' \mid \theta \in \text{Sym}(S \setminus \text{slots}(m_1 * c_1)), m_2' \in \text{Aut}(c_2)\} \neq$$

$$\min\{\theta * \min\{m_2 * m_2' \mid m_2' \in \text{Aut}(c_2)\} \mid \theta \in \text{Sym}(S \setminus \text{slots}(m_1 * c_1))\}$$

Open questions

- Is λ above the most generic possible, or are there examples of languages where the binders cannot be expressed this way?