Term Generalization for Idempotent Equational Theories

David M. Cerna and Temur Kutsia

Research Institute for Symbolic Computation (RISC)
Johannes Kepler University, Linz, Austria

June 1^{st}-3^{rd}, 2018

What is Term Generalization?

- Let us consider a signature of function and constant symbols $\Sigma = \{f_1, \cdots f_n, a_1, \cdots, a_m\}$ where each f_i has a positive arity. Together with a countably infinite set of variables $\mathcal{V}$ we can consider the language $\mathcal{L}$ constructible from $\Sigma \cup \mathcal{V}$.
- Now consider two arbitrary terms $t, s \in \mathcal{L}$. we can ask the following:
  - can we instantiate the variables of t and s using a substitution σ such that $t\sigma = s\sigma$ where $=$ is defined as syntactic equality? This is Unification.
  - can we find a term $r \in \mathcal{L}$ and substitutions σ_1 and σ_2 such that $r\sigma_1 = t$ and $r\sigma_2 = s$? This is Anti-Unification or Term Generalization, the problem we will focus on in this talk.
- Term generalization problems will be written $t \triangleq s$.

Generalization over an Equational Theory

- This is not a new problem, it was first introduced in G. Plotkin in 1970 and has been widely developed since then.
- Of most interest to the work discussed in this talk is the equational generalization problem.
- Rather than comparing terms over syntactic equality, one can check terms over particular equational theories.
- For example associativity, commutativity, unity, and, most importantly for us, idempotency.

Good Solutions?

- One might be asking themselves, at the moment, isn't it trivial to find such an r?
- Well...yes.
- Let $r = x$ where x is a variable, and let $\sigma_1 = \{x \mapsto t\}$ and $\sigma_2 = \{x \mapsto s\}$.
- If $t = f(t')$ and $s = f(s')$ then there is obviously a less general solution that is $r = f(x)$ where x is a variable, and let $\sigma_1 = \{x \mapsto t'\}$ and $\sigma_2 = \{x \mapsto s'\}$
- How do we formulate this rigorously to get a good definition of the least general solution? Does it exists uniquely?

LGGs: The Cream of the Crop

- Anyone who knows unification is probably familiar with the concept of an MGU or most general unifier. An LGG is a dual concept for term generalization.
- Defining this concept requires an ordering on generalizers. Consider the problem $s \triangleq t$ and r a solution.
- Let r' be a another solution. We say r' is <u>less general than r</u> if there exists a substitution σ such that $r' = r\sigma$. We ignore variable renamings.
- Of course the <u>least general generalizer</u> is a generalizer r'' such that no such substitution exists.
- Is it unique? Can there be multiple incomparable LGGs?

The Many Choices of Equational Reasoning

- In most cases LGGs are unique for syntactic equality. Adding an equational theory changes this and the result is many, not necessarily finite, incomparable LGGs.
- let us consider the term signature $\{f(\cdot,\cdot), a, b, c, d\}$ where f is interpreted as commutative over our equational theory C_f, that is $f(x,y) =_{C_f} f(y,x)$. Now consider the anti-unification problem

$$f(f(a,b), f(b,c)) \triangleq f(f(b,d), f(a,c))$$

- what are it's LGGs?

The Many Choices of Equational Reasoning

- We claim that

$$f(f(a,x), f(b,y)) \quad f(f(b,x), f(c,y))$$

are both LGGs.

- Notice that we cannot transform one into the other by a substitution.
- The next question is whether we found all of the incomparable LGGs, that is the minimal complete set of LGGs.
- There are 4 other reasonable generalizations one can derive:

$$f(f(a,x), f(z,y)) \quad f(f(z,x), f(b,y))$$
$$f(f(b,x), f(z,y)) \quad f(f(z,x), f(c,y))$$

But they are not LGGs and are instances of the above two.

An Infinite Set of LGGs

- It turns out that the minimal complete set of LGGs for commutativity will always be finite.
- What does it take to get an infinite minimal complete set of LGGs?
- This is not easy to answer, but we will consider a theory which has this property, Idempotency.

$$f(x,x) = x$$

- The key being that not only can we go from $f(x,x) \to x$, we can also go from $x \to f(x,x)$.

Two Idempotent Function Symbols

- In "Generalisation de termes en theorie equationnelle. Cas associatif-commutatif" by L. Pottier, an example was given of a pair of terms constructed using two idempotent function symbols whose set of generalizations contains an infinite number of incomparable generalizations.
- Specifically, the terms are constructed using the signature $\{f(\cdot,\cdot), g(\cdot,\cdot), a, b\}$ where f and g are idempotent. We refer to the equational theory as $I_{\{f,g\}}$.

$$f(a, b) \triangleq g(a, b)$$

A Infinite Sequence of Generalizations

- The following terms generalize the anti-unification problem:

$$g(f(a,x),f(y,b)) \quad f(g(a,x),g(y,b))$$

$$g(f(a,x),f(y,b))\,\{x \mapsto a\ ,\ y \mapsto b\} =_{I_{\{f,g\}}} g(a,b)$$

$$g(f(a,x),f(y,b))\,\{x \mapsto b\ ,\ y \mapsto a\} =_{I_{\{f,g\}}} f(a,b)$$

- This is not a complete set, but enough for constructing an infinite incomparable sequence.

$$S_0 = g(f(a,x),f(y,b))$$

$$S_{n+1} = f(g(f(a,x),f(y,b)),g(S_n,f(g(a,x),g(y,b))))$$

$$f(g(f(a,x),f(y,b)),g(S_n,f(g(a,x),g(y,b)))) \neq_{I_{\{f,g\}}} f(g(f(a,x),f(y,b)),g(S_{n+1},f(g(a,x),g(y,b))))$$

Do we Really Need Two Function Symbols?

- If one idempotent function symbol turns out to be finitary, then the above result would imply that joining two finitary theories can result in an infinitary theory. Unstable behavior!
- But does the above example really need both f and g?
- Considering f and g to be functions it is easy to imagine an h such that $h(a, a, b) = f(a, b)$ and $h(b, a, b) = g(a, b)$.
- What if we apply this reasoning to our problem and look at the anti-unification problem:

$$h(a, h(a, b)) \triangleq h(b, h(a, b))$$

One Function, Infinite Solutions

- The following terms generalize the anti-unification problem:

$$h(h(x, h(x, b)), h(a, h(x, b))) \quad h(f(x, h(a, x)), h(h(x, b), h(a, b)))$$

$$h(h(x, h(x, b)), h(a, h(x, b)))\,\{x \mapsto a\} =_{I_{\{h\}}} h(a, h(a, b))$$

$$h(h(x, h(x, b)), h(a, h(x, b)))\,\{x \mapsto b\} =_{I_{\{h\}}} h(b, h(a, b))$$

- Notice that the solutions are in some sense simpler and thus more fundamental. Less variables.
- Using the Pottier construction we can produce the an infinite set of incomparable LGGs.

Infinite but Neither Minimal nor Complete

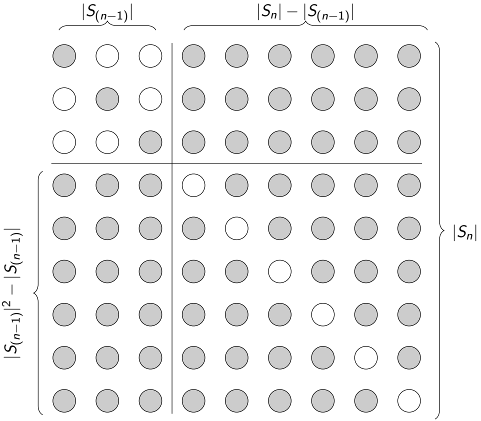
- Ok, we know that one idempotent function symbol is enough for infinitely many solutions, What is the simplest example resulting in infinitely many solutions?
- It turns out that
$$f(a, b) \triangleq f(b, a)$$
is enough
$$f(f(x_1, a), f(b, x_2)) \qquad f(f(x_1, b), f(a, x_2))$$
- This observation has lead to a generalization of the previous construction.

A minimal complete set

$$S_0 = \{f(f(x_1, a), f(b, x_2)), f(f(x_1, b), f(a, x_2))\}.$$
$$S_k = \{f(s_1, s_2) \mid s_1, s_2 \in S_{k-1}, s_1 \neq s_2\} \cup S_{k-1},\ \ k > 0.$$

- The limit S_∞ can be proven minimal complete for the above problem and bounds on the growth can be computed.

Proof: Growth of S_n is $O\left(2^{2^n}\right)$



$$|S_{(n+1)}| = |S_n|^2 - |S_{(n-1)}|^2 + |S_{(n-1)}| \qquad |S_1| = m \qquad |S_0| = 1$$

Is the S-hierarchy Enough?

- while the S-hierarchy works for $f(a,b) \triangleq f(b,a)$ it fails for slightly more complex problems, i.e.

$$f(a,f(a,b)) \triangleq f(a,f(b,a))$$

It captures an infinite number of incomparable generalizations, it also misses an infinite number because $f(a,b) \triangleq f(b,a)$ is embedded within this problem:

$$f(a,f(f(x,a),f(b,y))) \in S_\infty \qquad f(a,f(f(x,b),f(a,y))) \in S_\infty$$

$$f(f(a,f(f(x,a),f(b,y))),f(a,f(f(x,b),f(a,y)))) \in S_\infty$$

$$f(a,f(f(f(x,a),f(b,y)),f(f(x,b),f(a,y)))) \notin S_\infty$$

Capturing Minimal Completeness

- To construct a minimal complete set of LGGs we need to go beyond nesting generalizers. We are currently investigating approaches to this problem. For example, instead of computing generalizers:

$$f(a, f(f(x, a), f(b, y))) \quad f(a, f(f(x, a), f(b, y)))$$

  we construct a binding list

$$[x \mapsto f(a, x_2)]\,[x_2 \mapsto f(f(z_1, a), f(b, x_4))]$$
$$[x \mapsto f(a, x_2)]\,[x_2 \mapsto f(f(z_2, b), f(a, x_5))]$$

- using these binding list we can construct a binding tree:

  $[x \mapsto f(a, x_2)]\,[x_2 \mapsto f(f(z_2, b), f(a, x_5))]\,[x_2 \mapsto f(f(z_1, a), f(b, x_4))]$

  this tree can be extended by the node

$$[x_2 \mapsto f(f(f(z_1, a), f(b, x_4)), f(f(z_2, b), f(a, x_5)))]$$

  algorithmically.

Unfortunately, the Full Binding Tree is...

$[z \mapsto y]$

- $[y \mapsto x]$
  - $[x \mapsto f(f(a, f(a, x_1)), x_2)]$
    - $[x_2 \mapsto f(f(y_1, f(y_1, b)), f(a, f(a, y_2)))]$
    - $[x_2 \mapsto f(y_3, f(a, y_4))]$
    - $[x_2 \mapsto f(f(y_5, a), y_6)]$
      - $[y_6 \mapsto f(f(z_1, a), f(b, z_2))]$
      - $[y_6 \mapsto f(f(z_3, b), f(a, z_4))]$
      - $[y_6 \mapsto f(z_5, z_6)]$
  - $[x \mapsto f(f(a, f(x_4, a)), x_5)]$
    - $[x_5 \mapsto f(f(a, f(y_{11}, a)), f(y_{12}, f(b, y_{13})))]$
    - $[x_5 \mapsto f(a, f(b, y_{14}))]$
    - $[x_5 \mapsto f(f(a, y_{15}), y_{16})]$
      - $[y_{16} \mapsto f(f(z_7, a), f(b, z_8))]$
      - $[y_{16} \mapsto f(f(z_9, b), f(a, z_{10}))]$
      - $[y_{16} \mapsto f(z_{11}, z_{12})]$
- $[y \mapsto f(a, x_3)]$
  - $[x_3 \mapsto f(f(y_7, a), f(b, y_8))]$
  - $[x_3 \mapsto f(f(y_9, b), f(a, y_{10}))]$

This is the resulting binding tree when all binding list are computed and joined.

Notice that it is much more than the simple binding tree we discussed on the previous slide.

The complexity of these trees grows fast.

Expanding a Binding Tree

$[x \mapsto f(f(a, f(a, x_1)), x_2)]$
$[x_2 \mapsto f(f(y_1, f(y_1, b)), f(a, f(a, y_2)))]$
$[x_2 \mapsto f(y_3, f(a, y_4))]$
$[x_2 \mapsto f(f(y_5, a), y_6)]$
$[y_6 \mapsto f(f(z_1, a), f(b, z_2))]$
$[y_6 \mapsto f(f(z_3, b), f(a, z_4))]$
$[y_6 \mapsto f(z_5, z_6)]$

$[x \mapsto f(f(a, f(a, x_1)), x_2)]$
$[x_2 \mapsto f(f(y_1, f(y_1, b)), f(a, f(a, y_2)))]$
$[x_2 \mapsto f(y_3, f(a, y_4))]$
$[x_2 \mapsto f(f(y_5, a), y_6)]$
$[y_6 \mapsto f(f(z_1, a), f(b, z_2))]$
$[y_6 \mapsto f(f(z_3, b), f(a, z_4))]$
$[y_6 \mapsto f(z_5, z_6)]$
$[x_2 \mapsto f(f(w_1, f(a, w_2)), f(f(w_3, a), w_4))]$
$[w_4 \mapsto f(f(r_1, a), f(b, r_2))]$
$[w_4 \mapsto f(f(r_3, b), f(a, r_4))]$
$[w_4 \mapsto f(z_5, z_6)]$

- We conjecture that given the binding tree constructible from the term structure (without additional function symbols) then a minimal complete set of LGGs can be constructed by expansion.
- Proof: a work in progress

The End

Thank you for your time!