Type of Anti-Unification in Absorption Theories

Andrés Felipe González Barragan
Supervisors: Mauricio Ayala-Rincón (UnB)
Temur Kutsia(RISC - U. Linz)

Universidade de Brasília
Brasília, February 9 2023

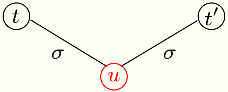
Outline

1. Motivation
2. Preliminaries
3. Absorption Theory
4. Analysis in Nonlinear Problems
5. Conclusions and Future work

Motivation

Unification

Finding a substitution that identifies two expressions (terms).



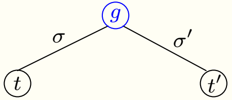
where $t\sigma \approx u \approx t'\sigma$.

Example 1

Identify the terms $h(g(a), y)$ and $h(g(z), f(w))$. Using the substitution $\sigma = \{y \mapsto f(w), z \mapsto a\}$ the expressions *unify* to $h(g(a), b)$.

Anti-unification

Finding the commonalities between two expressions (terms).
An expression with this commonalities is called a *generalisation*.



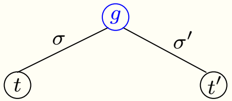
where $g\sigma \approx t$ and $g\sigma' \approx t'$.

Example 2

Generalize the terms $h(g(a), y)$ and $h(g(z), f(w))$.

Anti-unification

Finding the commonalities between two expressions (terms).
An expression with this commonalities is called a *generalisation*.

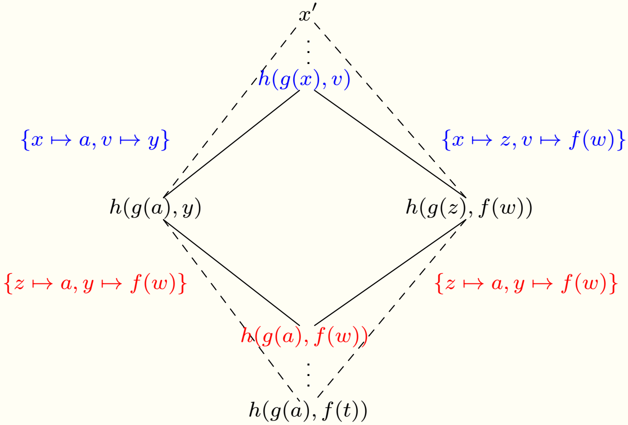


where $g\sigma \approx t$ and $g\sigma' \approx t'$.

Example 2

Generalize the terms $h(g(a), y)$ and $h(g(z), f(w))$.
generalisation: $h(g(x), v)$, with substitutions $\sigma = \{x \mapsto a, v \mapsto y\}$ and $\sigma' = \{x \mapsto z, v \mapsto f(w)\}$.

Unification and Anti-unification



Preliminaries

One interesting example of verbatim plagiarism:

- (Original sentence). All around the world, technology is continuing to become a part of everyday life, and its capabilities are progressing rapidly.
- (Possibly sentence with plagiarism). All over the world, technology has becomed a part of our lives, and its capabilities are progressing very quickly.

Preliminaries

Then finding the common parts and the differences in the sentences:

- All around the world, technology is continuing to become a part of everyday life, and its capabilities are progressing rapidly.
- All over the world, technology has becomed a part of our lives, and its capabilities are progressing very quickly.

All □ the world, technology □ a part of □, and its capabilities are progressing □.

Preliminaries

Applications of anti-unification include:

- searching parallel recursion schemes to transform sequential algorithms into parallel algorithms (Barwell et al. [BBH18]);
- preventing bugs and missconfigurations in software (Mehta et al. [MBK+20]);
- finding duplicate code and similarities;
- detecting code clones (i.e., plagiarism).

Preliminaries

- We consider an alphabet $\mathcal{A}$, that consists of Σ a signature with symbol functions and $\mathcal{V}$ a set of variables.
- A term construction over $\mathcal{A}$, denoted by $\mathcal{T}$, defined as usually:

$$t := x \mid f(t_1, \ldots, t_n)$$

- A finite set E that consists of equations $s \approx t$.
- A preorder relation $\preceq_E$, which states that $s \preceq_E t$ if there exists a substitution σ such that $s\sigma \approx_E t$.

Preliminaries

Anti-unification

- An anti-unification equation (AUE) between s and t is denoted by $s \overset{x}{\triangleq}_E t$, where x is called as label.
- One of our goals is to build a *minimal* complete set of generalisations ($mcsg_E(s,t)$):
  - Each $r \in mcsg_E(s,t)$ is an E-generalisation of s and t.
  - For each E-generalisation r of s and t, there exist $r' \in mcsg_E(s,t)$ such that $r \preceq_E r'$.
  - If $r, r' \in mcsg_E(s,t)$ and $r \preceq_E r'$ then $r =_\alpha r'$.

Preliminaries

Type of anti-unification problems

The type of an anti-unification modulo E problem is classified as below.

- *Nullary*(0): if there are terms s and t such that $\mathtt{mcsg}_E(s,t)$ does not exist. Also, called of *type zero*.
- *Unitary*(1): if for all s and t, $\mathtt{mcsg}_E(s,t)$ has just one generalisation.
- *Finitary*(ω): if for all s and t, $\mathtt{mcsg}_E(s,t)$ is finite.
- *Infinitary*(∞): there are terms s and t such that $\mathtt{mcsg}_E(s,t)$ is infinite.

Classification in some Theories

Theory	Type	Authors and References	Procedure or Term
Syntactic (∅)	1	G. Plotkin [Plo70, Rey70]	Dec, Sol, Rec
Associativy (A)	ω	M. Alpuente et al. [AEEM14]	A-left, A-right
Commutativy (C)	ω	M. Alpuente et al. [AEEM14]	C
Unitary (U)	ω	D. Cerna [CK20a]	Start-C, Sat-C, M
Idempotency$_{\geq 1}$(I$_{\geq 1}$)	∞	D. Cerna and T. Kutsia [CK20a]	M, Id-left,Id-right
			Id-both (1,2,3)
Unital$_{\geq 2}$ (U$_2$)	0	D. Cerna and T. Kutsia [CK20b]	$e_f \triangleq e_g$
			$f(g(x, y), x)$

Classification in some Theories

Theory	Type	Authors and References	Procedure or Term
AC, ACU	ω	M. Alpuente et al. [AEEM14]	AC-left, AC-right
AU_2, CU_2, ACU_2	0	D. Cerna and T. Kutsia [CK20b]	$e_f \triangleq e_g$ $f(g(x,y),x)$
$(\mathrm{UI})_2$, $(\mathrm{ACUI})_2$	0	D. Cerna and T. Kutsia [CK20b]	$e_f \triangleq e_g$ $f(g(f(x,y),e_f),x)$
Semirings (S), SC	0	D. Cerna [Cer20]	$e_f \triangleq e_g$ $\prod_{i=1}^{n} x$

Classification in some Theories

Theory	Type	Authors and References	Procedure or Term
Absorption$_{\geq 1}$ (Abs)	?	—	—
$(\mathrm{ACU})_2$, $(\mathrm{ACU})_2\mathtt{Abs}$	0	D. Cerna [Cer20]	$e_f \triangleq e_g$ $\prod_{i=1}^{n} x$
Simply-typed λ-calculus	0	D. Cerna and M. Buran[BC22]	$\lambda xy.f(x) \triangleq \lambda xy.f(y)$
$\mathrm{I}\mathtt{Abs}$,$(\mathrm{UI})_2\mathtt{Abs}$	$\emptyset, \infty$?	—	—

Collapse Theories

Collapse and Subterm Collapsing Theories

A theory E is a **collapse theory** iff there exists an axiom $A \in E$ of the form $t = x$, where t is a non-variable term and x is a variable. A theory E is called **subterm collapsing** iff one side of the equation is a proper subterm of the other.

For example:

- Unital: $f(x, \varepsilon_f) = x = f(\varepsilon_f, x)$;
- Idempotency: $f(x, x) = x$;
- Absorption: $f(\varepsilon_f, x) = \varepsilon_f = f(x, \varepsilon_f)$;

Collapse Theories

Absorption Theory

One important algebraic property in some magma is the absorbing property i.e. for some symbol function $f(x, \varepsilon_f) = \varepsilon_f$ or/and $f(\varepsilon_f, x) = \varepsilon_f$, we can find this property in semirings, rings, and in boolean algebras.

Example 3

Let's find the `lggs` of the AUE $\varepsilon_f \triangleq_{\texttt{Abs}} f(f(a, b), c)$.

Collapse Theories

Lemma 1 (Equivalence between E-generalisation and $\emptyset$-generalisation)

Let s and t E-normal forms, r is an E-generalisation of s and t if and only if r is an $\emptyset$-generalisation of s' and t' for some $s' \in [s]_E$ and $t' \in [t]_E$.

Example 3

Consider $\varepsilon_f \triangleq_{\texttt{Abs}} f(f(a,b),c)$. We can select the next terms in the classes:

$$f(f(a,b),c) \in [f(f(a,b),c)]_{\texttt{Abs}} \text{ and } f(f(\varepsilon_f,b),c) \in [\varepsilon_f]_{\texttt{Abs}}$$

Notice that the term $f(f(x,b),c)$ is a generalisation of $\varepsilon_f \triangleq_\emptyset f(\varepsilon_f,b),c)$, then is an Abs-generalisation too.

Absorption Theory

Algorithm for absorbing theory

To build the algorithm we consider a triple $\mathcal{C} := \langle A; S; \theta \rangle$ as a *configuration* in each step of the procedure, where:

- A is a set of anti-unification equations (AUEs);
- S is the *store*, the set of solved AUEs;
- θ is a substitution mapping the labels of the AUEs to their respective generalisations.

We always consider that all the terms in A are in normal form.

Absorption Theory

Inference Rules

Then we define the next rules

(Dec): **Decompose**

$$\langle \{ f(s_1, \ldots, s_n) \overset{x}{\triangleq} f(t_1, \ldots, t_n) \} \cup A; S; \theta \rangle$$

$$\overset{Dec}{\Longrightarrow} \langle \{ s_1 \overset{y_1}{\triangleq} t_1, \ldots, s_n \overset{y_n}{\triangleq} t_n \} \cup A; S; \theta \{ x \mapsto f(y_1, \ldots, y_n) \} \rangle$$

For f any function symbol, $n > 0$, and $y_1, \ldots, y_n$ are fresh variables.

Absorption Theory

Inference Rules

(Solve): **Solve**

$$\langle \{s \overset{x}{\triangleq} t\} \cup A; S; \theta \rangle \overset{Sol}{\Longrightarrow} \langle A; \{s \overset{x}{\triangleq} t\} \cup S; \theta \rangle$$

Where $head(s) \neq head(t)$ and some of them is a free symbol or they are non related absorption symbols.

Absorption Theory

Inference Rules

(ExpAL): **Expansion for Absorption, Left**

$$\langle\{\varepsilon_f \overset{x}{\triangleq} f(t_1,t_2)\}\cup A;S;\theta\rangle \overset{ExpAL}{\Longrightarrow} \langle\{\varepsilon_f \overset{y}{\triangleq} t_1\}\cup A;S;\theta\{x\mapsto f(y,t_2)\}\rangle$$

(ExpAR): **Expansion for Absorption, Right**

$$\langle\{\varepsilon_f \overset{x}{\triangleq} f(t_1,t_2)\}\cup A;S;\theta\rangle \overset{ExpAR}{\Longrightarrow} \langle\{\varepsilon_f \overset{y}{\triangleq} t_2\}\cup A;S;\theta\{x\mapsto f(t_1,y)\}\rangle$$

Where f is an absorption function symbol, and y is a fresh variable.

Absorption Theory

Inference Rules

(Mer): **Merge**

$$\langle \emptyset; \{s \overset{x}{\triangleq} t\} \cup \{s \overset{y}{\triangleq} t\} \cup S; \theta\rangle \overset{Mer}{\Longrightarrow} \langle \emptyset; \{s \overset{y}{\triangleq} t\} \cup S; \theta\{x \mapsto y\}\rangle$$

(Elim): **Eliminate**

$$\langle \{s \overset{x}{\triangleq} s\} \cup A; S; \theta\rangle \overset{Elim}{\Longrightarrow} \langle A; S; \theta\{x \mapsto s\}\rangle$$

Analysis in nonlinear problems

Analysis in nonlinear problems

Consider the AUE $\varepsilon_f \triangleq f(a, h(a))$. The `lggs` of this problem are given by

$$f(x, h(a)), f(a, x), \text{ and } f(x, h(x))$$

But with the last rules we just can computes only the first two `lggs`, and they are incomparable with $f(x, h(x))$.

Set of f nested positions

Let $pos(t)$ be the set of positions of the term t and $\sqsubseteq$ the prefix order over the positions.

$$oc_f(t) = \{q \mid \textit{head}(t\mid_q) \neq f \text{ and for all } p \sqsubset q, \textit{head}(t\mid_p) = f\}$$

Example 4

We determine the f-nested position of the next term:

$$s = f(f(a, f(h(b), f(a, h(b)))), g(c))$$

Then, $oc_f(s) = \{2, 11, 121, 1221, 1222\}$.

Absorption Theory

Inference Rules

(ExpAB): **Expansion for Absorption, Both**

$$\langle \{\varepsilon_f \overset{x}{\triangleq} f(t_1, t_2)\} \cup A; S; \theta \rangle \overset{ExpAL}{\Longrightarrow} \langle \{\varepsilon_f \overset{y_1}{\triangleq} t_1, \varepsilon_f \overset{y_2}{\triangleq} t_2\} \cup A; S; \theta\{x \mapsto f(y_1, y_2)\} \rangle$$

Where f is an absorption function symbol, and y_1 and y_2 are fresh variables.

Procedure ANT_UNIF

Absorption in a signature with constants

Considering again the AUE $\varepsilon_f \triangleq f(a, h(a))$, we noticed that the last rule does not compute the `lgg` $f(x, h(x))$, then we need to work with $\Sigma = \{f, \varepsilon_f\} \cup \Sigma'$, where Σ' have just a constants symbols.
We can find all the possible generalisations for any s and t, replacing a fresh variable in each f-nested position and considering the combination of repetitions.

Procedure ANT_UNIF

Example 5

Therefore the lggs of $\varepsilon_f \triangleq_{\texttt{Abs}} f(f(a,b), f(a,c))$ in the f-nested positions

$$oc_f(f(f(a,b), f(a,c)) = \{11, 12, 21, 22\}$$

The solutions are given by:

- $f(f(a,b), f(y, h(c))[x]_{11} = f(f(x,b), f(a,c)),$
- $f(f(a,b), f(y, h(c))[x]_{12} = f(f(a,x), f(a,c),$
- $f(f(a,b), f(y, h(c))[x]_{21} = f(f(a,b), f(x,c)),$
- $f(f(a,b), f(y, h(c))[x]_{22} = f(f(a,b), f(a,x))$
- $f(f(a,b), f(y, h(c))[x]_{11}[x]_{21} = f(f(x,b), f(x,c))$

Analysis in nonlinear problems

Analysis in nonlinear problems general case

Consider the AUE $g(\varepsilon_f, a) \triangleq g(f(h(\varepsilon_f), a), \varepsilon_f)$. The lggs of this problem give us a infinite set of incomparable lggs.
This solutions are given for the generalisations $g(f(w, a), y)$ and $g(f(h(x), z), y)$, where x is generate by the next grammar:

- $Y(0) = t$ for $t \in \{z\} \cup \mathcal{T}_g$,
- $Y(1) = y$,
- $Y(k) = f(Y(i), Y(j))$, where $i + j = k - 1$ for $k \geq 2$.

Analysis in nonlinear problems

Example 6

Some `lggs` of $g(\varepsilon_f, a) \triangleq g(f(h(\varepsilon_f), a), \varepsilon_f)$:

- $g(f(h(f(z,y)),z),y) = g(f(h(f(Y(0),Y(1))),z),y) = g(f(h(Y(2)),z),y)$
- $g(f(h(f(z,f(z,y))),z),y) = g(f(h(Y(3)),z),y)$
- $g(f(h(f(b,f(g(a,a),y))),z),y) = g(f(h(Y(3)),z),y)$

Conclusions and Future work

Conclusions

- We determine a terminating procedure for the restriction of the problem just considering a signature with constants, producing a finite set of `lggs` for any s and t.
- We find an AUE that generates an infinite set of incomparable `lggs` for the general problem, then the type of anti-unification for absorption theory is at least infinitary.

References

María Alpuente, Santiago Escobar, Javier Espert, and José Meseguer.
A modular order-sorted equational generalization algorithm.
Information and Computation, 235:98–136, 2014.

Adam D. Barwell, Christopher Brown, and Kevin Hammond.
Finding parallel functional pearls: Automatic parallel recursion scheme detection in haskell functions via anti-unification.
Future Gener. Comput. Syst., 79:669–686, 2018.

Michal Buran and David M. Cerna.
One or nothing: Anti-unification over the simply-typed lambda calculus.
CoRR, abs/2207.08918, 2022.

David M. Cerna.
Anti-unification and the theory of semirings.

Thanks!