Anti-unification and Generalization: A Survey

David M. Cerna and Temur Kutsia

JOHANNES KEPLER
UNIVERSITY LINZ

August 19^{th}-25^{th} 2022
IJCAI-23

Czech Academy
of Sciences

A Survey on Anti-unification

WHY?

An Unfamiliar Concept

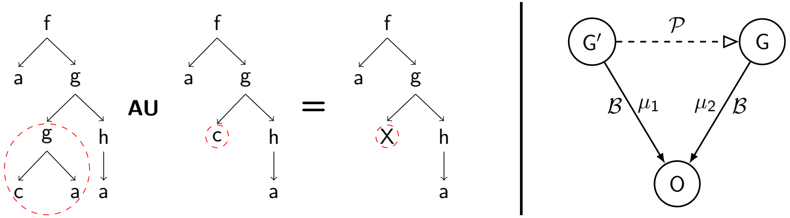
- For some (possibly many) members of the audience
  - *Anti-unification* is a **new** concept.
- Some may have heard of θ-subsumption.
  - Applications within **Inductive Logic Programming**
- We expect few are aware of the following:
  - Anti-unification is the dual operation to **unification**.
  - There exists **anti-unification algorithms** modulo various equational theories, over Higher-order languages, and within a variety of other settings.
  - Anti-unification has applications within **Formal Reasoning**, **Inductive Synthesis**, **Theory exploration**, **Program Analysis**, and **Program Repair**.

The Need For a Comprehensive Source

- Anti-unification remains obscure even though it is a useful and inexpensive technique for generalization and abstraction.
- Consider the recent work:

  Babble: Learning Better Abstractions with E-Graphs and Anti-Unification (2023), POPL, Cao et al.
- The authors provide a novel program compression mechanism using equational anti-unification.
- Their system is competitive with statistical learning based approaches, i.e. **Dreamcoder** [Ellis *et al.*, 2021].
- However the authors only address **Plotkin's foundational work**. This is likely due to the fractured nature of the literature and how time-intensive act of processing.

What is Anti-unification (AU)? (General Perspective)



- **Goal:** from $O_1, O_2 \in \mathcal{O}$ (symbolic expressions) derive $G \in \mathcal{O}$ possessing certain commonalities shared by O_1 and O_2.
- **Specification:** define (a) a class of mappings $\mathcal{M}$ from $\mathcal{O} \to \mathcal{O}$, (b) a base relation $\mathcal{B}$ consistent with $\mathcal{M}$, and (c) a preference relation $\mathcal{P}$ consistent with $\mathcal{B}$.
- **Result:** G is a $\mathcal{B}$-generalization of O_1 and O_2 and most $\mathcal{P}$-preferred ("better" than G').

Complete Sets and Types (A General Perspective)

- A set $\mathcal{G} \subset \mathcal{O}$ is called $\mathcal{P}$-complete set of $\mathcal{B}$-generalizations of $O_1, O_2 \in \mathcal{O}$ if:
  - **Soundness:** Every $G \in \mathcal{G}$ is a $\mathcal{B}$-generalization of O_1 and O_2.
  - **Completeness:** For each $\mathcal{B}$-generalization G' of O_1 and O_2, there exists $G \in \mathcal{G}$ such that $\mathcal{P}(G, G')$ (G is more preferred).
- Furthermore, $\mathcal{G}$ is minimal if:
  - **Minimality:** No distinct elements of $\mathcal{G}$ are $\mathcal{P}$-comparable: if $G_1, G_2 \in \mathcal{G}$ and $\mathcal{P}(G_1, G_2)$, then $G_1 = G_2$.
- Minimal Complete sets come in four Types:
  - **Unitary** $(\mathbf{1})$: $\mathcal{G}$ is a singleton,
  - **Finitary** (ω): $\mathcal{G}$ is finite and contains at least two elements,
  - **Infinitary** (∞): $\mathcal{G}$ is infinite,
  - **Nullary** $(\mathbf{0})$: $\mathcal{G}$ does not exist (minimality and completeness contradict each other).
- Types are extendable to generalization problems.

First-order Syntactic Generalization

Generic	Concrete
$\mathcal{O}$	The set of first-order terms
$\mathcal{M}$	First-order substitutions
$\mathcal{B}$	$\doteq$ (syntactic equality)
$\mathcal{P}$	$\succeq$ (more specific, less general): $s \succeq t$ iff $s \doteq t\sigma$ for some $\sigma \in \mathcal{M}$
$\equiv_{\mathcal{P}}$	Equi-generality: $\succeq$ and $\preceq$
Type	Unitary
Alg.	[Huet, 1976; Plotkin, 1970: Reynolds, 1970]

- Extendable to *first-order clausal generalization* and *relative θ-subsumption.*
- Clausal generalization is a special case of **equational generalization** (ACUI).

Least-Common Subsumer (Description Logics)

Generic	Concrete
$\mathcal{O}$	Concept descriptions
$\mathcal{M}$	Contains only the identity mapping
$\mathcal{B}$	$\sqsupseteq$ - for $C, D \in \mathcal{O}$, $C^{\mathcal{I}} \supseteq D^{\mathcal{I}}$ for all interpretations $\mathcal{I}$
$\mathcal{P}$	$\sqsubseteq$ - for $C, D \in \mathcal{O}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all interpretations $\mathcal{I}$
$\equiv_{\mathcal{P}}$	$\equiv$: $\sqsubseteq$ and $\sqsupseteq$
Type	Unitary for all four DLs
Alg.	[Baader *et al.*, 1999] for $\mathcal{EL}$, $\mathcal{FLE}$, $\mathcal{ALE}$, [Küsters and Molitor, 2001] for $\mathcal{ALEN}$

- Generalization (least common subsumer) in Description logics $\mathcal{EL}$, $\mathcal{FLE}$, $\mathcal{ALE}$, and $\mathcal{ALEN}$.

Open Problems and Conclusions

- We present a general framework for discussing and characterising anti-unification and generalization problems.
- Read this survey if you want to know:
  - What it means to **generalize** in various logic-based languages.
  - What are the **techniques** to compute such generalizations.
  - What are the **applications** of these methods and algorithms.

- There are many open problems including:
  - **Combining** anti-unification algorithms for disjoint (equational) theories.
  - Studying the influence of the **preference relation** on the **type and solution set** of generalization problems.
  - Studying **computational complexity and optimizations** of existing generalization problems.

Acknowledgements

Supported by Czech Science Foundation Grant **No. 22-06414L**, Austrian Science Fund project **P 35530**, and European Research Network on Formal Proofs COST action **CA20111** (**EuroProofNet**).

FUNDED BY

GAČR

CZECH SCIENCE FOUNDATION

FWF Austrian Science Fund

cost

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY