

Enumerative Type-Guided Program Synthesis for Higher-Order Polymorphic Language

DARYA VERZHBINSKY, University of California, San Diego

DANIEL WANG, University of California, San Diego

1 INTRODUCTION

Consider the task of implementing a function that has the following description: “If I give you a function, apply it to x . Otherwise, return x unchanged.” We can translate this function description into a Haskell type signature as follows:

$$f : \text{Maybe}(a \rightarrow a) \rightarrow x : a \rightarrow a \quad (1)$$

where f represents a possible function (hence the *Maybe* type), and x is of the same type that f takes in and returns. A possible solution to this problem is:

$$\backslash f \ x \rightarrow \text{fromMaybe } (\backslash y \rightarrow y) \ f \ x \quad (2)$$

which either extracts the function from the *Maybe* and applies it to x , or applies the identity function to x if the *Maybe* is *Nothing* (which is the same as returning x unchanged). As a component based program, it is idiomatic and concise, and doesn’t rely on pattern matching. It is potentially non-intuitive, though, and could be difficult for programmers to find on their own.

We present **TODONAME**, a tool that allows users to provide a type signature and optional examples, and get back a code snippet that consists of a composition of Haskell library functions. This allows users to speed up their programming by using our tool to synthesize small Haskell programs instead of doing it themselves.

2 RELATED WORK

The problem we are exploring is polymorphic type-guided synthesis via enumerative search. Related problems have been explored before.

MYTH. Myth [link] already explores enumerative search for type-directed synthesis, but it does not support polymorphism or typeclasses like we’d like our tool to.

TYGAR. TYGAR [link] uses Petri nets to tackle type-guided Haskell program synthesis, and is remarkable in that it also supports polymorphic types, typeclasses, and can check user-provided input-output examples. Because of its use of Petri nets, however, it has to reduce everything to first order, meaning that it can’t make use of lambdas, which limits the space of programs that TYGAR can pull from.

SYNQUID. Synquid [link] makes use of an enumerative search approach, lambdas, and polymorphism. Its type system, however, restricts type variables from unifying with arrow types. This prevents, for example, $\text{Maybe } a \rightarrow a$ and $\alpha_1 \rightarrow \alpha_0 \rightarrow b$ from unifying, even though in theory they could if $\text{Maybe } a$ unified with α_1 , and a unified with $\alpha_0 \rightarrow b$. This leads to some programs being unreachable and, in the context of higher-order components, limits the search space.

3 APPROACH

3.1 Synthesis Problem

Our synthesis *problem* consists of a component library, a query type, and optional input-output examples. A *solution* to the synthesis problem is a Haskell code snippet that has the desired type and works for the optionally given examples.

3.2 Extending Synquid’s enumeration strategy

Our approach builds upon Synquid, which solves its synthesis task by performing a top-down enumerative search and taking advantage of refinement types that are included in the initial type query.

Synquid synthesizes function applications in the following way. Given a goal $?? :: T$ that is an application, split it into two goals for the function and argument: $(?? :: \alpha \rightarrow T)$ and $(?? :: \alpha)$. Repeating this process in the first goal lets us use functions that take in more arguments:

$$\begin{aligned} & (?? :: \alpha_0 \rightarrow T)(?? :: \alpha_0) \\ & (?? :: \alpha_1 \rightarrow \alpha_0 \rightarrow T)(?? :: \alpha_1)(?? :: \alpha_0) \\ & (?? :: \alpha_2 \rightarrow \alpha_1 \rightarrow \alpha_0 \rightarrow T)(?? :: \alpha_2)(?? :: \alpha_1)(?? :: \alpha_0) \end{aligned}$$

This could repeat forever. To ensure search terminates, the components decide the cutoff point: if the longest arrow type in the components is $a \rightarrow b \rightarrow c \rightarrow d$, then we stop splitting the goal once we see $\alpha_2 \rightarrow \alpha_1 \rightarrow \alpha_0 \rightarrow T$.

However, this approach fails when type variables are instantiated with arrow types. One concrete example is the query $(a \rightarrow b, a) \rightarrow b$ with components `[fst, snd]`. The solution is `fst p (snd p)`, which uses `fst` as a two-argument function, which means `fst` must have filled the goal $\alpha_1 \rightarrow \alpha_0 \rightarrow T$. However, `fst` and `snd` are one-argument functions – the cutoff is $\alpha_0 \rightarrow T$, so we never end up considering the goal $\alpha_1 \rightarrow \alpha_0 \rightarrow T$.

Our solution is to use iterative deepening on the size of programs, rather than depth first search. Because we only enumerate programs up to a certain size anyways, we can be rid of the decision of “where to stop splitting the goal”.

We also replace the measure of program size with “number of components” rather than “recursion depth”, because size tends to generate deeper programs earlier, and deep programs like `f (g (h x))` tend to be more programmer-relevant than flat programs.

3.3 Memoization (caching subproblems during enumeration)

Because of its enumerative and iterative deepening qualities, our tool generates many repeated subproblems over the course of the search, wasting a lot of time redoing searches it has already done. This makes memoization crucial to practical top-down enumeration, which Myth [link] was able to solve in a first-order setting. In our polymorphic context, however, memoization is non-trivial. With polymorphism, goal types can contain free type variables. This causes a couple issues:

- (1) **Type unification clashes:** With free type variables as goals, each cached program could cause the type variable to unify with a different type. For example, if the goal is $\alpha_0 \rightarrow Int$, two possible solutions are: `length`, which has type $\forall \alpha. [\alpha] \rightarrow Int$, resulting in α_0 and $[\alpha]$ being unified, and `incr`, which has type $Int \rightarrow Int$, resulting in α_0 and Int being unified. How do we ensure that this won’t cause any clashes when retrieving the problems from the memo map when later trying to synthesize component arguments?

Table 1. TYGAR vs. Our tool

N	Name	Query	Time (s): TYGAR	Time (s): Ours	Speedup
1	appBoth	$(a \rightarrow b) \rightarrow (a \rightarrow c) \rightarrow a \rightarrow (b, c)$	8.5	7.5	1.137333333
2	test	$\text{Bool} \rightarrow a \rightarrow \text{Maybe } a$	8.0	6.0	1.337815126
3	both	$(a \rightarrow b) \rightarrow (a, a) \rightarrow (b, b)$	4.1	-	-
4	firstJust	$a \rightarrow [\text{Maybe } a] \rightarrow a$	18.4	6.5	2.824615385
5	mapEither	$(a \rightarrow \text{Either } b \ c) \rightarrow [a] \rightarrow ([b], [c])$	4.8	3.2	1.49375
6	mapMaybes	$(a \rightarrow \text{Maybe } b) \rightarrow [a] \rightarrow \text{Maybe } b$	8.0	5.2	1.544230769
7	mergeEither	$\text{Either } a \ (\text{Either } a \ b) \rightarrow \text{Either } a \ b$	3.8	181.5	0.02077707357
8	mbToEither	$\text{Maybe } a \rightarrow b \rightarrow \text{Either } a \ b$	12.7	4.2	3.026128266
9	cartProduct	$[a] \rightarrow [b] \rightarrow [[(a, b)]]$	-	-	-
10	multiAppPair	$(a \rightarrow b, a \rightarrow c) \rightarrow a \rightarrow (b, c)$	14.0	-	-
11	map	$(a \rightarrow b) \rightarrow [a] \rightarrow [b]$	2.1	3.7	0.5564516129
12	repl-funcs	$(a \rightarrow b) \rightarrow \text{Int} \rightarrow [a \rightarrow b]$	1.1	1.2	0.9411764706
13	mbAppFirst	$b \rightarrow (a \rightarrow b) \rightarrow [a] \rightarrow b$	3.2	9.5	0.3396226415
14	2partApp	$(a \rightarrow b) \rightarrow (b \rightarrow c) \rightarrow [a] \rightarrow [c]$	3.4	10.6	0.3163841808
15	resolveEither	$\text{Either } a \ b \rightarrow (a \rightarrow b) \rightarrow b$	4.0	10.3	0.3912621359
16	dedupe	$\text{Eq } a \Rightarrow [a] \rightarrow [a]$	19.0	-	-
17	multiApp	$(a \rightarrow b \rightarrow c) \rightarrow (a \rightarrow b) \rightarrow a \rightarrow c$	6.9	9.6	0.7165271967
18	singleList	$a \rightarrow [a]$	7.7	4.4	1.750566893
19	head-last	$[a] \rightarrow (a, a)$	257.3	12.0	21.4093178
20	head-rest	$[a] \rightarrow (a, [a])$	1.5	3.3	0.4693251534

- (2) **Redundancy:** At different points in the search, the free variables will have different names. This means two different goals, $\alpha_0 \rightarrow \text{Int}$ and $\alpha_1 \rightarrow \text{Int}$ would be syntactically different but semantically equivalent. How do we avoid storing the same code snippets in multiple locations?

To resolve the first issue, we rename any incoming free type variables to be "fresh" before using them as the map lookup key. That way, if two different goals with free variables come in, say $\alpha_0 \rightarrow \text{Int}$ and $\alpha_1 \rightarrow \text{Int}$, they will each be renamed to $\beta_0 \rightarrow \text{Int}$ and $\beta_0 \rightarrow \text{Int}$, meaning that they will hit the same entry in the map, despite having different original names.

To resolve the second issue, we use type inference to infer the type of the retrieved programs. For example, if we retrieve a program `length`, Similarly, if we retrieve a program `incr`,

4 RESULTS

We tested our results on 20 queries, that were based on the queries from TYGAR's [link] paper and λ^2 [link], and compared our tool with TYGAR, which is the tool that most resembles our own. The results can be found in Table 1.

Analysis. TYGAR did better than our tool in (TODO) out of the 20 tests. For most of the tests, the difference between the 2 tools is quite small. There are additionally some tests that we do much better in (TODO list them) and some tests that we do much worse in (TODO list them). Test # (TODO) shows how we are able to synthesize lambdas whereas TYGAR cannot.

These results are quite encouraging. For one, since our tool takes advantage of lambdas and TYGAR doesn't, our tool is more expressive and therefore searches through more programs, yet is still competitive. We also have only implemented the most basic form of memoization, and plan to improve upon it to make our tool even faster (see [Future Work](#)).

Quality of Results. For the most part, both TYGAR and our tool returned the same programs. If they didn't return the same program, the programs were equivalent, so the differences between the results of the two tools is insignificant.

5 FUTURE WORK

Our memoization tool is still very basic and we think there are multiple ways in which we can improve upon what we have:

- (1) Re-organize our memo map to store programs first based on size, and then find programs based on query. This would make lookup much faster.
- (2) Take advantage of sub-typing in the memo keys so that the same programs aren't stored multiple times. For example, all programs in goal $Int \rightarrow Int$ should be in goal $\alpha_0 \rightarrow Int$. When we lookup $\alpha_0 \rightarrow Int$, we could first look at $Int \rightarrow Int$ and then move on to other programs that are more general.

REFERENCES

TODO