1 Introduction

One of the significant problems in using Prolog is the lack of static typing. Prolog inherited this feature from predicate calculus, where it roots lie. In the foundational study of propositions and quantification, types are often omitted because it is said they can already be expressed. For example, instead of saying $\forall x: \text{nat}. A(x)$ we can say $\forall x. \text{nat}(x) \supset A(x)$ if we have a predicate nat that expresses the type nat. Similarly, we can express $\exists x: \text{nat}. A(x)$ as $\exists x. \text{nat}(x) \land A(x)$. Predicates that provide an extensional representation of types are not difficult to come by. For example, we can define (and have defined) the natural numbers with two constructors $\text{z}$ and $\text{s}$ and the rules

\[
\begin{array}{c}
\text{nat}(\text{z}) \text{ true} \\
\text{nat}(\text{s}(N)) \text{ true}
\end{array}
\]

\[
\begin{array}{c}
\text{nat}_z \\
\text{nat}_s
\end{array}
\]

Foundationally, this approach may have some merit, but it also has some problems. One is that propositions such as $\forall x: \text{nat}. \text{append}(x, \text{nil}, x)$ which are meaningless become either true or false when written in an untyped way: $\forall x. \text{nat}(x) \supset \text{append}(x, \text{nil}, x)$. In a language like Prolog this has dire consequences because we compute with intuitively meaningless propositions and bogus proofs, leading to unexpected behavior. A second problem is that the untyped approach does not extend well to higher-order logic, where we want to quantify over propositions and not just data. In fact, several times in history well-regarded researchers such as Frege or Church have attempted to avoid the organizing principles of types, leading to inconsistent logics.

In this lecture we explore the question if we may still be able to use the idea of defining types via (unary) predicates and obtain something we can statically check and that executes efficiently at the same time. The answer is “yes”, and the lessons learned from
this has also had some impact on functional programming in the guise of refinement
types [FP91, DP03, Dav97].

There have been multiple approaches to types in the logic programming community
(see [Pfe92] for various articles and technical realizations). We will not go into a specific
decidable language of types, although much of what we show in this lecture applies to
several systems that are different in their respective technical details.

2 Modes and Types

Let’s reconsider something simple like addition on unary natural numbers.

\[
\begin{align*}
\text{nat}(z) & \quad \text{nat}_z \\
\text{nat}(N) & \quad \text{nat}_s \\
\text{plus}(z, N, N) & \quad \text{pz} \\
\text{plus}(M, N, P) & \quad \text{ps}
\end{align*}
\]

Now we want to show the combined *mode and type specification:*

\[
\text{plus}(+\text{nat}, +\text{nat}, -\text{nat})
\]

which we interpret as follows: if proof search is initiated with a goal \(\text{plus}(m, n, P)\) where \(\text{nat}(m)\) and \(\text{nat}(n)\) and proof search succeeds, then \(P = p\) with \(\text{nat}(p)\).

Rigorously, we would have to prove this by induction over the structure of computation (that is, proof search). In the absence of such an operational semantics, we prove it by induction over the structure of the rules. Assume we are searching for a proof of \(\text{plus}(m, n, P)\) for a variable \(P\) and terms \(m\) and \(n\) with \(\text{nat}(m)\) and \(\text{nat}(n)\).

**Case:** Rule \(pz\). We know \(\text{nat}(z)\) (which adds no new information because \(\text{nat}_z\) already knew this) and \(\text{nat}(n)\). Applying the rule will succeed and instantiate \(P = n\) and so \(\text{nat}(P)\).

**Case:** Rule \(ps\). We know \(m = s(m')\) and \(\text{nat}(s(m'))\) and also \(\text{nat}(n)\). From the first fact, by inversion (only rule \(\text{nat}_s\) could be used to prove this) we obtain \(\text{nat}(m')\). Now we can appeal to the induction hypothesis: if the subgoal \(\text{plus}(m', n, P')\) succeeds, then \(P' = p'\) for some term \(p'\) and \(\text{nat}(p')\). Then \(\text{nat}(s(p'))\) by rule \(\text{nat}_s\).

So far, there is not much new or interesting in this when compared to types as we know them from functional languages. But we can define new and interesting types as predicates and reason about them in the same style. For example, we can distinguish the even and odd numbers and reason about the properties of addition.

\[
\begin{align*}
\text{even}(z) & \quad \text{ev}_z \\
\text{odd}(N) & \quad \text{ev}_s \\
\text{even}(N) & \quad \text{od}_s
\end{align*}
\]

---

1Don’t confuse the + in \(\text{plus}(+\text{nat}, +\text{nat}, -\text{nat})\) with plus for addition. It refers to the mode where that argument is given as input natural number while – is the mode where that argument is computed as a natural number result.
Let's try to check that adding two even numbers results in an even number.

\[
\begin{array}{c}
\text{plus}(z, N, N) \quad \text{pz} \\
\text{plus}(M, N, P) \quad \text{ps} \\
\text{plus}(+\text{even}, +\text{even}, -\text{even})
\end{array}
\]

**Case:** Rule pz.

\[
\begin{array}{l}
even(z) \\
even(N) \\
even(N)
\end{array}
\]

mode +even of first arg.

mode +even of second arg.

previous line

**Case:** Rule ps.

\[
\begin{array}{l}
even(s(M)) \\
odd(M) \\
even(N)
\end{array}
\]

mode +even of first arg.

by inversion from previous line

mode +even of second arg.

At this point we are stuck because we cannot apply the induction hypothesis, only knowing that \( M \) is odd.

So we need to generalize our declaration to the pair

\[
\begin{array}{l}
(i) \quad \text{plus}(+\text{even}, +\text{even}, -\text{even}) \\
(ii) \quad \text{plus}(+\text{odd}, +\text{even}, -\text{odd})
\end{array}
\]

and restart our proof whose beginning is unchanged.

**Case (i):** Rule pz.

\[
\begin{array}{l}
even(z) \\
even(N) \\
even(N)
\end{array}
\]

mode +even of first arg.

mode +even of second arg.

previous line

**Case (i):** Rule ps.

\[
\begin{array}{l}
even(s(M)) \\
odd(M) \\
even(N) \\
odd(P) \\
even(s(P))
\end{array}
\]

mode +even of first arg.

by inversion from previous line

mode +even of second arg.

by i.h.(ii)

by rule evs

**Case (ii):** Rule pz.

**CONSTLog Lecture Notes**

FRANK PFENNING, ANDRÉ PLATZER
L16.4 Types as Predicates

odd(z)  mode +odd of first arg.
Contradiction  by inversion (no rule concluding odd(z))

Case (ii): Rule ps.

odd(s(M))  mode +odd of first arg.,
even(M)  by inversion from previous line
even(N)  mode +even of second arg.,
even(P)  by i.h.(i)
odd(s(P))  by rule od_

Note there that the case px in the proof of (ii) is impossible: the rule px cannot apply if the first argument of plus is odd. From the contradiction in this case we can infer anything, in particular that the third argument will be odd if the search succeeds, which it never will.

We see two differences here already to a system of types for functional languages such as ML: types as predicates have a natural notion of multiple related types (such as the mutually recursive even and odd numbers, as well as arbitrary natural numbers), and a given predicate such as plus may have multiple types, all of them necessary for type-checking purposes. We have to check for even and odd simultaneously.

3 Subtyping

Types defined as predicates come with a natural notion of subtyping. For two predicates s and t we write s ≤ t if ∀x. s(x) ⊃ t(x), that is, every element satisfying s satisfies t.

To appreciate the need for subtyping, we consider once again binary numbers starting with least-significant bit and numbers in standard form (no leading zeros). We defined this slightly differently from last lecture by stipulating that in a term b0(N), N must be positive. This enforces that it cannot be e, which represents zero and is therefore not positive.

\[
\frac{\text{std}(e)}{\text{std}(e)} \quad \frac{\text{pos}(N)}{\text{std}(b0(N))} \quad \frac{\text{std}(N)}{\text{std}(b1(N))} \quad \frac{\text{std}(1)}{\text{std}(1)}
\]

We now recall the increment predicate and try to verify that, if given a standard number it will construct a positive one.

\[
\frac{\text{inc}(e, b1(e))}{\text{inc}(e, b1(e))} \quad \frac{\text{inc}(b0(M), b1(M))}{\text{inc}(b0(M), b1(M))} \quad \frac{\text{inc}(M, N)}{\text{inc}(b1(M), b0(N))} \quad \frac{\text{inc}(1)}{\text{inc}(1)}
\]

\[
\text{inc}(+, \text{std}, -\text{pos})
\]
Case: Rule $\text{inc}_e$.

$$\text{pos}(b1(e))$$

Case: Rule $\text{inc}_0$.

$$\text{std}(b0(M))$$  
$$\text{pos}(M)$$  
$$\text{std}(M)$$  
$$\text{pos}(b1(M))$$

Case: Rule $\text{inc}_1$.

$$\text{std}(b1(M))$$  
$$\text{std}(M)$$  
$$\text{pos}(N)$$  
$$\text{pos}(b0(N))$$

At this point the proof is complete, if we can show that $\text{pos} \leq \text{std}$. This is now a property that no longer requires appeal to the definition of predicate $\text{inc}$; it is just a property of the two types. We can proceed by induction (actually, just a proof by cases is required) on the definition of $\text{pos}$.

Case: Rule $\text{pos}_0$.

$$\text{pos}(N)$$  
$$\text{std}(b0(N))$$

Case: Rule $\text{pos}_1$.

$$\text{std}(N)$$  
$$\text{std}(b0(N))$$

Next we see how this kind of static type checking (phrased here as theorem proving) can help uncover errors. For example, we may want to check that

$$\text{inc}(-\text{std}, +\text{std})$$

Case: Rule $\text{inc}_e$. Then $\text{std}(e)$.

Case: Rule $\text{inc}_0$.

$$\text{std}(b1(N))$$  
$$\text{std}(N)$$  
$$\text{Need: } \text{pos}(N)$$  
$$\text{std}(b0(N))$$

CONSTLOG Lecture Notes

FRANK PFENNING, ANDRÉ PLATZER
There is no way to fix the missing step in the second case (we didn’t even get around to the third case). \( \text{std}(N) \) does not imply \( \text{pos}(N) \), with \( N = e \) as a counterexample. Indeed, one solution for

\[
?- \text{inc}(M, b1(e))
\]

is \( M = b0(e) \) which is not in standard form.

At this point we might consider some other properties. Let’s define some new types, such as \( \text{zero}(N) \), and \( \text{empty}(N) \) which never succeeds:

\[
\begin{array}{cc}
\text{zero}(e) & \text{zero}_e \\
\hline
\text{no rule for empty}(N)
\end{array}
\]

Now we can show, with type checking that a query \( \text{inc}(M, e) \) cannot succeed. The type we ascribe is

\[
\text{inc}(\text{-empty}, +\text{zero})
\]

which expresses that if a query \( \text{inc}(M, n) \) with \( \text{zero}(n) \) succeeds with \( M = m \), then \( \text{empty}(m) \). Since there is no such \( m \), this means if \( \text{inc} \) has the given type then decrementing \( \text{zero} \) can not succeed. This means it either doesn’t terminate or it fails after a finite number of steps.

Now to the type checking:

**Case:** Rule \( \text{inc}_e \).

\[
\begin{array}{ll}
\text{zero}(b1(e)) & \text{second argument} \\
\text{Contradiction} & \text{by inversion (no rule concludes zero}(b1(e)))
\end{array}
\]

**Case:** Rule \( \text{inc}_0 \).

\[
\begin{array}{ll}
\text{zero}(b1(M)) & \text{second argument} \\
\text{Contradiction} & \text{by inversion}
\end{array}
\]

**Case:** Rule \( \text{inc}_1 \).

\[
\begin{array}{ll}
\text{zero}(b0(M)) & \text{second argument} \\
\text{Contradiction} & \text{by inversion}
\end{array}
\]

All cases are impossible, so the type \( \text{inc}(\text{-empty}, +\text{zero}) \) is correct.

As a last example we revisited the even/odd distinction, now on binary numbers. We could just look at the least significant bit, but we arrange it such that \( \text{even} \leq \text{std} \) and \( \text{odd} \leq \text{pos} \) to ease working with these types.

\[
\begin{array}{cc}
\text{pos}(N) & \text{ev}_0 \\
\text{even}(b0(N)) & \text{std}(N) \\
\text{odd}(b1(N)) & \text{od}_1
\end{array}
\]

We leave it to the reader to now verify that

\[
\text{inc}(+\text{even}, -\text{odd}) \\
\text{inc}(+\text{odd}, -\text{even})
\]
4 Refinement types for functional languages

The idea that we have more precise types than just nat (like even and odd) or binary numbers (like std, pos, zero, empty) could be a priori useful for functional languages as well, not just for logic programming.

The main complication is that we also need to include intersection types [CDCV81, Rey91] to make this work. We retain the usual data types, but we add data sort declarations that declare refinements [FP91]. The examples in this section and many more can be found in a conservative extension of Standard ML with datasort refinements called Cidre [Dav97], available on GitHub².

For example, we can define even and odd unary numbers as follows:

```plaintext
datatype nat = z | s of nat

datasort even = z | s of odd
  and odd = s of even
```

But now if we have a simple function such as

```plaintext
fun succ x = s(x)
```

we find that it has multiple function types at once:

```plaintext
succ : (nat -> nat) & (even -> odd) & (odd -> even)
```

So we need to be able to ascribe multiple types to a function or expression. This is what the intersection type operator $A \cap B$ achieves, which we write as $A \& B$ in concrete syntax. Sometimes, several types are needed. For example

```plaintext
fun twice f x = f (f x)
```

then we should be able to show (among many other types)

```plaintext
twice : ((nat -> nat) -> nat -> nat)
  & (((even -> odd) & (odd -> even)) -> (even -> even))
  & (((even -> odd) & (odd -> even)) -> (odd -> odd))
```

The resulting system has some remarkable properties, such as decidability of type inference, bidirectional type checking, and conservative extension over ML. It is implemented in the Cidre front end, which accepts the full syntax of Standard ML and uses stylized comments to assign refinement types that are then checked.

You probably have seen one example where this might have been helpful. For example, for propositions with implication and conjunction, one can define proof terms via the following data type

²https://github.com/rowandavies/sml-cidre
datatype term = Fun of var * term
  | Pair of term * term
  | Var of var
  | App of term * term
  | Fst of term
  | Snd of term

In a way, this was a compromise, since we distinguished, in the problem statement and the algorithm, between checkable and synthesizing terms. The corresponding data type declaration would be something like

datatype cterm = Fun of var * cterm
  | Pair of cterm * cterm
  | Syn of sterm

and sterm = Var of var
  | App of sterm * cterm
  | Fst of sterm
  | Snd of sterm

but there are two drawbacks: (1) we need to make the transition from synthesizing to checkable terms explicit (see Syn constructor), which complicates practical examples, and (2) now everywhere that terms are used, even in places where the distinction would be insignificant, we have to be cognizant and specific about whether we are working with checkable or synthesizing terms. With refinement types, we would first declare the type of terms, and then think of checkable and synthesizing terms as refinements.

 datatype term = Fun of var * term
  | Pair of term * term
  | Var of var
  | App of term * term
  | Fst of term
  | Snd of term

data sort cterm = Fun of var * cterm
  | Pair of cterm * cterm
  | sterm

and sterm = Var of var
  | App of sterm * cterm
  | Fst of sterm
  | Snd of sterm

We can now freely use either term (where it doesn’t matter) or cterm or sterm where the distinction is significant.

References


