The complexity of type inference for System F

Perry Hart

UMN PL Seminar

March 29, 2023

Overview

Let $\mathbb{T}$ denote a type system.

Type inference: Given a raw term t of $\mathbb{T}$, decide whether t is typable.

Theorem

1. *Type inference for STLC is decidable in polynomial time.*
2. *Type inference for Core ML is decidable in $O(2^n)$ time and is* EXPTIME*-hard under logspace reduction.*

The more expressive the language, the harder the decision problem.

Example

Deciding whether a propositional sentence is true is PTIME-complete.

Deciding whether $\exists x.\varphi(x)$ with $\varphi(x)$ a propositional formula is NP-complete.

Theorem (Henglein and Mairson)

Type inference for System F is EXPTIME*-hard under logspace reduction.*

Corollary

Type inference for System F requires exponential time.

Proof strategy

For each integer $k \geq 1$, find a logspace reduction

$$\text{total TM } M \text{ and string } x \text{ of length } n \quad \mapsto \quad \text{raw term } \Psi_k(M, x),$$

where M rejects x in 2^{n^k} steps if and only if $\Psi_k(M, x)$ is non-typable.

Such a reduction is known for Core ML.

Difficulty

The non-typable term $\Psi_k(M, x)$ in Core ML could be typable in System F.

Construct a term $A_k(M, x)$ of System F in logspace such that

$$A_k(M, x) \to^* \textbf{false} \text{ when } M \text{ accepts } x \text{ in } 2^{n^k} \text{ steps}$$
$$A_k(M, x) \to^* \textbf{true} \text{ when } M \text{ rejects } x \text{ in } 2^{n^k} \text{ steps.}$$

Take

$$\Psi_k(M, x) \;:=\; (\Lambda\alpha.\lambda(x:\alpha).xx)\,(A_k(M, x)\,(\Lambda\alpha.\lambda(x:\alpha).x)\,(\Lambda\alpha.\lambda(y:\alpha).yy))$$

Construct a term $A_k(M,x)$ of System F in logspace such that

$$A_k(M,x) \to^* \textbf{false} \text{ when } M \text{ accepts } x \text{ in } 2^{n^k} \text{ steps}$$
$$A_k(M,x) \to^* \textbf{true} \text{ when } M \text{ rejects } x \text{ in } 2^{n^k} \text{ steps.}$$

Take

$$\Psi_k(M,x) \;:=\; (\Lambda\alpha.\lambda(x:\alpha).xx)\,(A_k(M,x)\,(\Lambda\alpha.\lambda(x:\alpha).x)\,(\Lambda\alpha.\lambda(y:\alpha).yy))$$

If $A_k(M,x) \to^* \textbf{false}$, then

$$\Psi_k(M,x) \;\longrightarrow^*\; \underbrace{(\Lambda\alpha.\lambda(x:\alpha).xx)\,(\Lambda\alpha.\lambda(y:\alpha).yy)}_{\text{non-terminating}}.$$

If $A_k(M,x) \to^* \textbf{true}$, then we must show that $\Psi_k(M,x)$ is typable.

Encoding Turing machines with System F

Let M be a TM.

The n-th configuration $\mathrm{ID}_{M,x}(n)$ of M on input x is the triple (q, L, R) where, after exactly n steps,

- q is the state of M,
- L is the (finite) list of symbols to the left of the read head, and
- R the list of symbols at or to the right of the read head.

We can find a λ-term δ_M in logspace such that

$$\delta_M \ \overline{\mathsf{ID}_{M,x}(n)} \quad \longrightarrow^* \quad \overline{\mathsf{ID}_{M,x}(n+1)}$$

for all $n \in \mathbb{N}$.

We can find a λ-term δ_M in logspace such that

$$\delta_M \ \overline{\mathsf{ID}_{M,x}(n)} \quad \longrightarrow^* \quad \overline{\mathsf{ID}_{M,x}(n+1)}$$

for all $n \in \mathbb{N}$.

Moreover, the term

$$\left(\overline{2^{n^k}}[T_{\mathsf{ID}_{M,x}}]\ \delta_M\right) \overline{\mathsf{ID}_{M,x}(0)}$$

has normal form $\overline{\mathsf{ID}_{M,x}(2^{n^k})}$. If $(q, L, R) = \mathsf{ID}_{M,x}(2^{n^k})$, then take

$$A_k(M,x) \ := \ \overline{q}[\mathrm{Bool}]\ \mathit{false} \cdots \mathit{false}\ \underbrace{\mathit{true} \cdots \mathit{true}}_{|Q_F| \text{ copies}}.$$

If M accepts x in 2^k steps, then $A_k(M, x)$ reduces to **true**.

Therefore,

$$(\lambda(x:\tau).x[\tau]x)\,(A_k(M,x)[\tau][\tau \to \tau]\,(\Lambda\alpha.\lambda(x:\alpha).x)\,(\lambda(y:\tau).y[\tau]y)) \qquad (\tau := \forall\alpha.\alpha \to \alpha)$$

has type τ, so that $\Psi_k(M, x)$ is typable.

If M accepts x in 2^k steps, then $A_k(M,x)$ reduces to **true**.

Therefore,

$$(\lambda(x:\tau).x[\tau]x)\,(A_k(M,x)[\tau][\tau\to\tau]\,(\Lambda\alpha.\lambda(x:\alpha).x)\,(\lambda(y:\tau).y[\tau]y))$$
$$(\tau := \forall\alpha.\alpha\to\alpha)$$

has type τ, so that $\Psi_k(M,x)$ is typable.

Since $A_k(M,x)$ can be constructed in logspace, so can $\Psi_k(M,x)$.

Related questions

- The decidability of type inference for System F was an open question until 2012, then answered in the negative by Fujita and Schubert.
- Pfenning proved that partial type inference for System F is undecidable.
- The set of all strongly normalizing terms of System F is RE-complete (hence undecidable).