Sistemas de Tipos

Cristiano Damiani Vasconcellos

cristiano.vasconcellos@udesc.br

Departamento de Ciência da Computação
Universidade do Estado de Santa Catarina

Tipos

Tipos: Coleção de valores ou objetos que possuem alguma propriedade em comum.

Na matemática, tipos impõe restrições que evitam paradoxos. Universos não tipados apresentam inconsistências lógicas tais como o **paradoxo de Russell**.

Paradoxo de Russell

Paradoxo de Russell

Alguns conjuntos não são membros de si próprios, como por exemplo o conjunto de todas as cadeiras. Outros, como por exemplo o conjunto formado por tudo que não é cadeira, são membros de si mesmos. Definindo R como o conjunto de todos os conjuntos que não são membros de si próprio:

$$R = \{A \mid A \notin A\}$$

- Se R é membro dele mesmo então, por definição, R não é membro de R.
- Se R não é membro de R então, por definição, R é membro de R.

O próprio Russell respondeu seu paradoxo usando a teoria de tipos, definindo uma hierarquia para as proposições. Um dado predicado é valido para todos objetos que estiverem em um mesmo nível (ou forem do mesmo tipo).

Tipos

Linguagens que não definem um intervalo de valores que uma variável pode armazenar são classificadas como não tipadas. Essas linguagens suportam um único tipo que representa todos os valores. λ-cálculo é um exemplo extremo de linguagens não tipadas.

Uma linguagem de programação é considerada segura (*safe*) se todos os erros de tipos podem ser detectados, ou seja, os tipos não podem ser violados. Linguagens não tipadas podem ser consideradas seguras efetuando a verificação em tempo de execução.

Tipos

A definição de um sistema de tipos no projeto de linguagens de programação é útil para:

- **Estruturação dos programas e documentação:** os tipo representam abstrações dos dados manipulados pelo programa e podem ajudar na compreensão do código.
- **Detecção de erros:** uma grande variedades de erros podem ser detectados automaticamente quando dados e funções são usados de forma inconsistente.
- **Eficiência:** informações sobre tipos permitem ao computador executar otimizações no código gerado.

Sistemas de Tipos

Sistemas de tipos são conjuntos de regras de inferência que permitem atribuir tipos as variáveis e expressões de linguagens de programação. O principal objetivo de um sistema de tipos é determinar, em tempo de compilação, se um programa é bem comportado, garantindo a ausência de erros de tipos em tempo de execução. Um sistema capaz de fornecer essa garantia é dito consistente (*sound*).

Para que seja possível provar a consistência do sistema de tipos é necessária a sua formalização.

Prova Matemática: Verificação de uma proposição por encadeamento de deduções lógicas a partir de um conjunto de axiomas.

Sistemas de Tipos

A definição formal de um sistema de tipos é feita por um conjunto de enunciados (regras) denominadas sentenças (*judgments*). Sentenças são afirmações sobre objetos sintáticos de um determinado tipo. Uma sentença tem a forma:

$$\Gamma \vdash e : \sigma$$

Essa sentença é lida como: no contexto Γ a expressão e tem tipo σ ou Γ implica (deriva) em e ter tipo σ. Sendo Γ um contexto, possivelmente vazio, onde estão definidos os tipos das variáveis que ocorrem livres em e ($\Gamma = \{x_1 : \sigma_1, x_2 : \sigma_2, \ldots, x_n : \sigma_n\}$).

A forma geral das regras de inferência é:

$$\frac{\Gamma_1 \vdash e_1 : \sigma_1 \quad \Gamma_2 \vdash e_2 : \sigma_2 \quad \ldots \quad \Gamma_n \vdash e_n : \sigma_n}{\Gamma \vdash e : \sigma}$$

As sentenças acima da linha horizontal são as premissas e a sentença abaixo é a conclusão. Por exemplo:

$$\frac{\Gamma \vdash e_1 : \mathit{Nat} \quad \Gamma \vdash e_2 : \mathit{Nat}}{\Gamma \vdash e_1 + e_2 : \mathit{Nat}}$$

λ-Cálculo Simplesmente Tipado

Variaveis de Tipos	α, β
Variáveis de Expressões	x, y, z
Expressões e	$::= x \mid \lambda x.\, e \mid e\ e'$
Tipo Simples α	$::= \tau \mid \tau \to \tau'$

$$\Gamma \vdash x : \tau \ (VAR) \quad \{x : \tau\} \in \Gamma$$

$$\frac{\Gamma \vdash e : \tau \to \tau' \quad \Gamma \vdash e' : \tau}{\Gamma \vdash e\ e' : \tau'}\ (APP) \qquad \frac{\Gamma, x : \tau' \vdash e : \tau}{\Gamma \vdash \lambda x.e : \tau' \to \tau}\ (ABS)$$

$\Gamma, x : \tau$ representada $\Gamma \cup \{x : \tau\}$, sendo que Γ não apresenta qualquer suposição de tipo para x.

Unificação

Unificação é a ideia central do processo de inferência de tipos, um unificador para dois tipos é uma substituição S que: $S\tau_1 = S\tau_2$.

Uma **substituição** é uma função que mapeia variáveis de tipos em expressões de tipos. Uma substituição pode ser representada como: $S = \{\alpha_1 \mapsto \tau_1, \alpha_2 \mapsto \tau_2, \ldots, \alpha_n \mapsto \tau_n\}$. A aplicação de uma substituição S em um tipo τ $(S\tau)$ resulta na troca de todas as variáveis de tipo que ocorrem em τ e pertencem ao domínio de S pelo tipo correspondente em S.

A composição de substituições é representada por $S \circ S'$. Um unificador S_g é chamado de **unificador mais geral** se, para qualquer outro unificador S, existe uma substituição S' tal que $S' \circ S_g = S$.

Inferência de Tipos

```
data SimpleType = TVar Id
                |TArr SimpleType SimpleType
                deriving (Eq, Show)

data Expr       = Var Id
                |App Expr Expr
                |Lam Id Expr
                deriving (Eq, Show)
```

Inferência de Tipos

```
tiExpr g (Var i)   = return (tiContext g i, [])
tiExpr g (App e e')=
         do  (t, s1) <- tiExpr g e
             (t', s2) <- tiExpr g e'
             b <- freshVar
             let s3 = unify (apply s2 t) (t' --> b)
             return (apply s3 b, s3 @@ s2 @@ s1)
tiExpr g (Lam i e) =
         do  b <- freshVar
             (t, s) <- tiExpr (g /+/ [i:>:b]) e
             return (apply s (b --> t), s)
```

Tipo Produto (*Product Type, Record*)

O produto de dois tipos consistem em um par ordenado de valores, sendo cada valor do tipo especificado. Podemos generalizar o tipo produto como o produto de um conjunto finito de n tipos, sendo $n \geq 0$. O tipo **unit** (tipo unitário) é representado pelo produto nulo.

$$\Gamma \vdash \langle\rangle : \text{unit}$$

$$\frac{\Gamma \vdash e : \tau \quad \Gamma \vdash e' : \tau'}{\Gamma \vdash e \times e' : \tau \times \tau'} \ (PAIR)$$

$$\frac{\Gamma \vdash e \times e : \tau \times \tau'}{\Gamma \vdash \pi_1(e \times e') : \tau} \ (PROJ) \ \frac{\Gamma \vdash \ e \times e' : \tau \times \tau'}{\Gamma \vdash \pi_2(e \times e') : \tau'}$$

Tipo União Disjunta (*Disjoint Union, Sum Type, Tagged Union*)

A união disjunta de dois tipos oferece uma escolha entre dois elementos de tipos possivelmente distintos. Cada um dos tipos é marcado com uma etiqueta (construtor do tipo) que permite a seleção por casamento de padrões (*pattern match*).

$$\frac{\Gamma \vdash e : \tau}{\Gamma \vdash e + e' : \tau + \tau'}\ (SUM)\ \frac{\Gamma \vdash e' : \tau'}{\Gamma \vdash e + e' : \tau + \tau'}$$

$$\frac{\Gamma, x_1 : \tau_1 \vdash e_1 : \tau \quad \Gamma, x_2 : \tau_2 \vdash e_2 : \tau \quad \Gamma \vdash e : \tau_1 + \tau_2}{\Gamma \vdash \text{case } e \text{ of } \{x_1 \Rightarrow e_1 \mid x_2 \Rightarrow e_2\} : \tau}\ (CASE)$$

Booleanos

Um exemplo simples do tipo união disjunta é o tipo booleano:

$$\Gamma \vdash \mathit{True} : \mathit{Bool} \qquad\qquad \Gamma \vdash \mathit{False} : \mathit{Bool}$$

$$\frac{\Gamma \vdash e : \mathit{Bool} \quad \Gamma \vdash e_1 : \tau \quad \Gamma \vdash e_2 : \tau}{\Gamma \vdash \text{if } e \text{ then } e_1 \text{ else } e_2 : \tau} \ (\mathit{IF})$$

Tipos Recursivos

Muitas linguagens permitem a definição de tipos recursivos, para podermos representar tipos recursivos em usamos um **operador de ponto fixo de tipos**. A expressão de tipo $\mu\alpha.\tau$ denota o isomorfismo dos tipos que satisfazem a equação $\mu\alpha.\tau \cong \{\alpha \mapsto \mu\alpha.\tau\}\tau$.

Por exemplo, o tipo Lista de inteiros pode ser definido como:

$$\mu\alpha.\langle\rangle + (\mathsf{Int} \times \alpha)$$

Que admite infinitas substituições de α por $\langle\rangle + (\mathsf{Int} \times \alpha)$:

$$\begin{array}{l}
\langle\rangle + (\mathtt{Int} \times \alpha) \\
\langle\rangle + (\mathtt{Int} \times \langle\rangle + (\mathtt{Int} \times \alpha)) \\
\langle\rangle + (\mathtt{Int} \times \langle\rangle + (\mathtt{Int} \times \langle\rangle + (\mathtt{Int} \times \alpha))) \\
\langle\rangle + (\mathtt{Int} \times \langle\rangle + (\mathtt{Int} \times \langle\rangle + (\mathtt{Int} \times \langle\rangle + (\mathtt{Int} \times \alpha)))) \ \ldots
\end{array}$$

Correspondência Curry-Howard

Demonstra uma correspondência direta entre tipos e teoremas. Uma função é uma prova, e o tipo de uma função é a fórmula provada.

Essa correspondência é demostrada com a **lógica intuicionista**.

Lógica Clássica

Três princípios fundamentais:

- **Reflexividade** – toda proposição deriva de si mesma: $\varphi \vdash \varphi$
- **Terceiro-excluído** – toda proposição é verdadeira ou falsa: $\vdash \varphi \vee \neg\varphi$
- **Não-Contradição** – não é possível que uma proposição seja simultaneamente verdadeira e falsa: $\vdash \neg(\varphi \wedge \neg\varphi)$

Lógica Intuicionista

Um sistema lógico mais fraco que a lógica clássica, possui um número menor de teoremas que podem ser demonstrados. Algo somente é verdade caso exista uma **prova construtiva**, portanto provas por absurdo não são permitidas.

Nesse sistema não existe **princípio do terceiro excluído**. Para uma prova construtiva do terceiro excluído, seria necessária uma prova da validade ou da falsidade de cada possível fórmula proposicional, o que é impossível.

A negação de uma proposição $\neg\varphi$ é definida como a obtenção do falso caso φ for vista como verdade, simbolicamente:

$$\neg\varphi \equiv \varphi \rightarrow \bot$$

Dedução Natural Intuicionista

$$\Gamma, \varphi \vdash \varphi \ (Ax)$$

$$\frac{\Gamma \vdash \varphi \quad \Gamma \vdash \psi}{\Gamma \vdash \varphi \wedge \psi} \ (\wedge I) \qquad \frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \varphi} \ (\wedge E) \frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \psi}$$

$$\frac{\Gamma \vdash \varphi}{\Gamma \vdash \varphi \vee \psi} \ (\vee I) \frac{\Gamma \vdash \psi}{\Gamma \vdash \varphi \vee \psi} \qquad \frac{\Gamma, \varphi \vdash \rho \quad \Gamma, \psi \vdash \rho \quad \Gamma \vdash \varphi \vee \psi}{\Gamma \vdash \rho} \ (\vee E)$$

$$\frac{\Gamma, \varphi \vdash \psi}{\Gamma \vdash \varphi \to \psi} \ (\to I) \qquad \frac{\Gamma \vdash \varphi \to \psi \quad \Gamma \vdash \varphi}{\Gamma \vdash \psi} \ (\to E)$$

$$\frac{\Gamma \vdash \bot}{\Gamma \vdash \varphi} \ (\bot E)$$

Fórmulas **Não** Deriváveis

As seguintes fórmulas **NÃO** são deriváveis na lógica intuicionista

$$\varphi \vee \neg\varphi$$

$$\neg\neg\varphi \to \varphi$$

$$(\neg\psi \to \neg\varphi) \to (\varphi \to \psi)$$

$$((\varphi \to \psi) \to \varphi) \to \varphi$$

$$\neg(\varphi \wedge \psi) \to (\neg\varphi \vee \neg\psi)$$

Correspondência Curry–Howard

$$\Gamma, \varphi \vdash \varphi \ (Ax) \qquad \Gamma, x : \tau \vdash x : \tau \ (VAR)$$

$$\frac{\Gamma \vdash \varphi \to \psi \quad \Gamma \vdash \varphi}{\Gamma \vdash \psi} \ (\to E) \qquad \frac{\Gamma \vdash e : \tau \to \tau' \quad \Gamma \vdash e' : \tau}{\Gamma \vdash e\ e' : \tau'} \ (APP)$$

$$\frac{\Gamma, \varphi \vdash \psi}{\Gamma \vdash \varphi \to \psi} \ (\to I) \qquad \frac{\Gamma, x : \tau' \vdash e : \tau}{\Gamma \vdash \lambda x.e : \tau' \to \tau} \ (ABS)$$

Correspondência Curry–Howard

$$\frac{\Gamma \vdash \varphi \quad \Gamma \vdash \psi}{\Gamma \vdash \varphi \wedge \psi}\ (\wedge I) \qquad \frac{\Gamma \vdash e : \tau \quad \Gamma \vdash e' : \tau'}{\Gamma \vdash e \times e' : \tau \times \tau'}\ (PROD)$$

$$\frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \varphi}\ (\wedge E) \qquad \frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \psi}$$

$$\frac{\Gamma \vdash e \times e : \tau \times \tau'}{\Gamma \vdash \pi_1(e \times e') : \tau}\ (PROJ) \qquad \frac{\Gamma \vdash \ e \times e' : \tau \times \tau'}{\Gamma \vdash \pi_2(e \times e') : \tau'}$$

Correspondência Curry–Howard

$$\frac{\Gamma \vdash \varphi}{\Gamma \vdash \varphi \vee \psi}\ (\vee I)\quad \frac{\Gamma \vdash \psi}{\Gamma \vdash \varphi \vee \psi}$$

$$\frac{\Gamma \vdash e : \tau}{\Gamma \vdash e + e' : \tau + \tau'}\ (SUM)\quad \frac{\Gamma \vdash e' : \tau'}{\Gamma \vdash e + e' : \tau + \tau'}$$

$$\frac{\Gamma, \varphi \vdash \rho \quad \Gamma, \psi \vdash \rho \quad \Gamma \vdash \varphi \vee \psi}{\Gamma \vdash \rho}\ (\vee E)$$

$$\frac{\Gamma, x_1 : \tau_1 \vdash e_1 : \tau \quad \Gamma, x_2 : \tau_2 \vdash e_2 : \tau \quad \Gamma \vdash e : \tau_1 + \tau_2}{\Gamma \vdash \text{case } e \text{ of } \{x_1 \Rightarrow e_1 \mid x_2 \Rightarrow e_2\} : \tau}\ (CASE)$$

Tipos Recursivos

Esse isomorfismo é definido por regras de conversão de um tipo $\mu\alpha.\tau$ em $[\mu\alpha.\tau/\alpha]\tau$ e vice-versa.

$$\frac{\Gamma \vdash e : \{\alpha \mapsto \mu\alpha.\tau\}\tau}{\Gamma \vdash \text{fold } e : \mu\alpha.\tau}\ (\textit{FOLD}) \qquad \frac{\Gamma \vdash e : \mu\alpha.\tau}{\Gamma \vdash \text{unfold } e : \{\alpha \mapsto \mu\alpha.\tau\}\tau}\ (\textit{UNFOLD})$$

Sistema Hindley-Milner

Em λ-cálculo simplesmente tipado os tipos são monomórficos as variáveis de tipos representam um único tipo em um contexto. No sistema *Hindley-Milner* são introduzidos tipos quantificados para o suporte ao **polimorfismo paramétrico**.

Variaveis de Tipos	α, β, γ		
Tipo Simples	τ	$::=$	$\alpha \mid \tau \to \tau'$
Tipo Polimórfico	σ	$::=$	$\tau \mid \forall \alpha.\sigma$

Figura: Expressões de Tipos

$$
\begin{aligned}
\mathit{ftv}(\alpha) &= \{\alpha\} \\
\mathit{ftv}(\tau \to \tau') &= \mathit{ftv}(\tau) \cup \mathit{ftv}(\tau') \\
\mathit{ftv}(\forall \alpha.\sigma) &= \mathit{ftv}(\sigma) - \{\alpha\}
\end{aligned}
$$

$\mathit{ftv}(\Gamma)$ denota a união $\mathit{ftv}(\sigma)$ para todo tipo σ que ocorre em Γ.

$$\frac{\beta_i \notin \mathit{ftv}(\forall\overline{\alpha}.\tau) \quad \tau' = \{\overline{\alpha} \mapsto \overline{\tau}\}\tau}{\forall\overline{\alpha}.\tau \leqslant \forall\overline{\beta}.\tau'}$$

Informalmente podemos dizer que o tipo $\forall\overline{\beta}.\tau$ é mais especifico ou o mesmo que $\forall\overline{\alpha}.\tau'$.

Variáveis		x, y, z
Expressões e	$::=$	x
		$\mid \lambda x.\, e$
		$\mid e\ e'$
		$\mid$ let $x = e$ in e'

Figura: Sintaxe de Expressões

Sistema Hindley-Milner

$$\Gamma \vdash x : \sigma \qquad \{x : \sigma\} \in \Gamma \qquad \text{(VAR)}$$

$$\frac{\Gamma \vdash e : \sigma}{\Gamma \vdash e : \sigma'} \qquad (\sigma \leqslant \sigma') \qquad \text{(INST)}$$

$$\frac{\Gamma \vdash e : \sigma}{\Gamma \vdash e : \forall \alpha.\sigma} \qquad (\alpha \notin \mathit{ftv}(\Gamma)) \qquad \text{(GEN)}$$

Figura: Sistema de tipos *Hindley-Milner*

Sistema Hindley-Milner

$$\frac{\Gamma \vdash e : \tau' \to \tau \qquad \Gamma \vdash e' : \tau'}{\Gamma \vdash e\ e' : \tau} \quad \text{(APP)}$$

$$\frac{\Gamma, x : \tau' \vdash e : \tau}{\Gamma \vdash \lambda x.e : \tau' \to \tau} \quad \text{(ABS)}$$

$$\frac{\Gamma \vdash e : \sigma \qquad \Gamma, x : \sigma \vdash e' : \tau}{\Gamma \vdash \text{let } x = e \text{ in } e' : \tau} \quad \text{(LET)}$$

Figura: Sistema de tipos *Hindley-Milner*

Algoritmo W

$W(\Gamma, x)$ =

Se $\Gamma(x) = \forall\alpha_1 ... \alpha_2.\tau$ então $(\{\alpha_i \mapsto \beta_i\}\tau, Id)$
senão *Falha*, sendo β_i *fresh*

$W(\Gamma, e\ e')$ =

let $(\tau, S_1) = W(\Gamma, e)$
$(\tau', S_2) = W(S_1\Gamma, e')$
S = *unificar* $(S_2\tau, \tau' \to \beta)$, sendo β *fresh*
in $(S\beta, S \circ S_2 \circ S_1)$

Algoritmo W

$$
\begin{aligned}
&W(\Gamma, \lambda x.e) \ = \\
&\qquad \texttt{let } (\tau, S) \ = \ W(\Gamma, x : \beta, e) \\
&\qquad \texttt{in } (S(\beta \to \tau), S)
\end{aligned}
$$

$$
\begin{aligned}
&W(\Gamma, \mathit{let}\ \ x = e\ \ \mathit{in}\ \ e') \ = \\
&\qquad \texttt{let } (\tau, S_1) \ = \ W(\Gamma, e) \\
&\qquad\qquad (\tau', S_2) \ = \ W(S_1\Gamma, x : \mathit{fechamento}\ \ (S_1\Gamma, \tau), e') \\
&\qquad \texttt{in } (\tau', S_1 \circ S_2)
\end{aligned}
$$

Sendo $\mathit{fechamento}(\Gamma, \tau) = \forall\overline{\alpha}.\tau$ e $\overline{\alpha} = \mathit{ftv}(\tau) - \mathit{ftv}(\Gamma)$.