Унификация посредством поиска путей с контекстно-свободными ограничениями в графе Source-tracking unification

Екатерина Вербицкая

Лаборатория языковых инструментов JetBrains

6 ноября 2020

TLDR

Задачу унификации можно свести к
поиску путей с КС ограничениями в графе[1]

[1] Choppella, V., and Haynes, C. T. (2005). Source-tracking unification.

План докалада

- Что такое унификация
- Как задача унификации представима в виде графа
- Какой язык будем использовать в качестве ограничений
- Почему это работает
- Какую дополнительную информацию можно получить из пути

Унификация

Даны два терма t, s

Задача: найти подстановку на свободных переменных термов (унификатор) θ, такую что

$$t\theta = s\theta$$

Подстановка

Терм: $\mathcal{T} :: \mathcal{V} \mid \mathcal{F}^n\ \mathcal{T}_1 \dots \mathcal{T}_n$

Подстановка: $\theta :: \mathcal{V} \to \mathcal{T}$

Применение подстановки $t\{x_1 \mapsto t_1, \dots, x_k \mapsto t_k\}$:
одновременно заменить свободные переменные x_i терма t на t_i

$$(f\ x\ a\ (g\ z)\ y)\{x \mapsto h\ a\ y, z \mapsto y\} = f\ (h\ a\ y)\ a\ (g\ y)\ y$$

Применение унификации

```
apply :: (a -> b) -> a -> b
apply f x = f x

f :: Int -> Int
f x = x + 1

apply_f :: ?
apply_f = apply f
```

Унифицируем `a -> b` и `Int -> Int`, получаем `a == Int`, `b == Int`

```
apply_f :: Int -> Int
```

Простой алгоритм унификации

Будем искать подстановку как множество уравнений $\mathcal{E} = \{t_i = s_i\}$

- Упрощение термов: $(f\ \ t_1 \dots t_n = g\ \ s_1 \dots s_m) \in \mathcal{E}$
  - Если f, g — различные константы, то $\mathcal{E} = \perp$
  - Иначе заменяем уравнение в $\mathcal{E}$ на множество $t_1 = s_1, \dots, t_n = s_n$
- Переориентация: $(t = x) \in \mathcal{E}$
  - Если t — терм, x — переменная, заменяем в $\mathcal{E}$ уравнение на $x = t$
- Элиминация переменных: $(x = t) \in \mathcal{E}$, x входит в какое-то уравнение
  - Если x входит в t, $t \equiv x$, то удаляем уравнение из $\mathcal{E}$
  - Иначе, если x входит в t, то $\mathcal{E} = \perp$
  - Иначе, подставляем t вместо x во всех уравнениях в $\mathcal{E}$

Унификация: пример

$$\{node\ El\ T\ T = node\ 1\ (node\ 2\ emp\ emp)\ (node\ 2\ emp\ emp)\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, T = node\ 2\ emp\ emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, node\ 2\ emp\ emp = node\ 2\ emp\ emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, 2 = 2, emp = emp, emp = emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp\}$$

Унификация: пример

$$\{node\ El\ T\ T = node\ 1\ (node\ 2\ emp\ emp)\ (node\ 3\ emp\ emp)\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, T = node\ 3\ emp\ emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, node\ 2\ emp\ emp = node\ 3\ emp\ emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, 2 = 3, emp = emp, emp = emp\}$$

$$\perp$$

Чем плох простой алгоритм

- Не очень эффективный
- Не говорит, почему унификация не завершилась успехом

Граф унификации