О сравнении Рефал-образцов

Антонина Непейвода

ИПС РАН

Рабочее совещание МГТУ им. Н. Э. Баумана и ИПС РАН по Рефалу
11 июня 2019

Постановка задачи

Автоматические преобразования программ на Рефале могут порождать предложения, недостижимые из-за поглощения их образцов образцами более приоритетных предложений.

Пример

```
F {  e.x1 'A' e.x2 = ... ;
     t.y1 t.y2 e.x1 = ...;
     e.x1 t.y1 e.x2 'B' = ...;
 }
```

Третье предложение лишнее — его образец является частным случаем образца `t.y1 t.y2 e.x1`.

Как удалять подобные предложения?

Определения

$\mathcal{V}_\mathcal{T}$ — множество переменных типа $\mathcal{T}$, $\mathcal{V} = \bigcup^\mathcal{T} \mathcal{V}_\mathcal{T}$.

Σ — *неограниченный* алфавит констант.

Определение

Плоский образец P *— строка в алфавите* $\mathcal{V} \cup \Sigma$. *Образец* P *линеен, если кратность каждой* e*-переменной в нем равна* 1. *Подстановка* σ *— морфизм из* $(\mathcal{V} \cup \Sigma)^*$ *в* $(\mathcal{V} \cup \Sigma)^*$, *сохраняющий константы (т.е. для всех* $\mathbf{A} \in \Sigma$ $\sigma(\mathbf{A}) = \mathbf{A}$*).*

Далее все образцы плоские и по умолчанию линейные.

Пример

Образец e.$\mathtt{x}$ **A** e.$\mathtt{x}$ плоский, но не линейный.

Образец s.$\mathtt{z}$ e.$\mathtt{x}_1$ e.$\mathtt{x}_2$ s.$\mathtt{z}$ — линейный и плоский.

Языки, распознаваемые образцами

Определение

Языком $\mathscr{L}(\mathsf{P})$*, распознаваемым образцом* P*, назовем множество элементов* $\Phi \in \Sigma^*$*, для которых существует подстановка* σ*:* $\sigma(\mathsf{P}) = \Phi$*.*
Образец P_1 *сводится к образцу* P_2*, если* $\mathscr{L}(\mathsf{P}_1) \subseteq \mathscr{L}(\mathsf{P}_2)$*.*

$\sigma(\mathtt{e.x}) = \varepsilon$ допустимо! В терминологии pattern languages — рассматриваются E-pattern languages (сокращение от Erasing Pattern Languages).

- Язык, распознаваемый образцом-строкой $\mathsf{P} \in \Sigma^*$, есть $\{\mathsf{P}\}$.
- Язык, распознаваемый образцом $\mathsf{P} = \mathtt{e.x}_1\,\mathtt{e.x}_2\,\ldots\,\mathtt{e.x}_\mathtt{n}$, есть все множество Σ^*.

Эквивалентность образцов

Если P_1, P_2 оба из $(\mathcal{V}_e \cup \Sigma)^*$ и линейны, то:

- P_1 сводится к P_2 $\Leftrightarrow$ существует подстановка σ такая, что $\sigma(\mathsf{P}_2) = \mathsf{P}_1$;
- вычислительная сложность проверки сводимости образца P_1 к образцу P_2 линейна от суммы длин P_1 и P_2.

Вопрос

Верно ли, что если $\mathscr{L}(\mathsf{P}_1) = \mathscr{L}(\mathsf{P}_2)$ (т.е. образцы эквивалентны), то существует переименовка переменных σ такая, что $\sigma(\mathsf{P}_1) = \mathsf{P}_2$?

Эквивалентность образцов

Вопрос

Пусть $\mathsf{P}_1, \mathsf{P}_2 \in (\mathcal{V}_{\mathsf{e}} \cup \Sigma)^*$ и линейны. Верно ли, что если $\mathscr{L}(\mathsf{P}_1) = \mathscr{L}(\mathsf{P}_2)$, то существует переименовка переменных σ такая, что $\sigma(\mathsf{P}_1) = \mathsf{P}_2$?

Ответ: неверно. Рассмотрим $\mathsf{P}_1 = \mathsf{e}.\mathtt{x}$, $\mathsf{P}_2 = \mathsf{e}.\mathtt{x}\,\mathsf{e}.\mathtt{y}$.

А если допустить только образцы без идущих подряд e-переменных?

Эквивалентность образцов

Определение

Дан образец P *и служебные символы* $\triangleright$ *и* $\triangleleft$. *Терминальным блоком называется подслово* α *образца* $\triangleright$ P $\triangleleft$ *такое, что:*

- α *не содержит* e-*переменных;*
- *вхождение* α *в* P *либо является префиксом, либо непосредственно следует за* e-*переменной;*
- *вхождение* α *в* P *либо является суффиксом, либо непосредственно предшествует* e-*переменной.*

В образце может быть больше чем один терминальный блок. Последовательность таких блоков, входящих в P, записываем через $\|$.

Пример

Образец $e.x_1$ **A B** $e.x_2$ $e.x_3$ **C** соответствует последовательности терминальных блоков $\triangleright \|$**A B**$\|$**C** $\triangleleft$.

Эквивалентность образцов

Вопрос

Пусть $P_1, P_2 \in (\mathcal{V}_e \cup \Sigma)^*$, линейны и не содержат идущих подряд e-переменных. Верно ли, что если $\mathscr{L}(P_1) = \mathscr{L}(P_2)$, то существует переименовка переменных σ такая, что $\sigma(P_1) = P_2$?

Ответ: да, верно.

Существуют подстановки σ_1, σ_2 такие, что $\sigma_1(P_1) = P_2$, $\sigma_2(P_2) = P_1$. Значит, последовательности терминальных блоков образцов P_1 и P_2 совпадают. Каждая такая последовательность с точностью до переименовки переменных соответствует единственному образцу без идущих подряд e-переменных.

Краткие и избыточные образцы

Определение

Образец P_1 *называется кратким, если любой образец* P_2 *такой, что* $\mathscr{L}(\mathrm{P}_1) = \mathscr{L}(\mathrm{P}_2)$, *имеет длину, не меньшую, чем* P_1.
Иначе P_1 *называется избыточным.*

Пример

Образец $\mathrm{e}.\mathtt{x}_1$ $\mathrm{e}.\mathtt{x}_2$ $\mathrm{e}.\mathtt{x}_1$ $\mathrm{e}.\mathtt{x}_2$ избыточен.

Образец $\mathrm{e}.\mathtt{x}_1$ $\mathrm{e}.\mathtt{x}_1$ является кратким.

Критерий избыточности образца (Reidenbach, 2004)

Образец P избыточен, если существует представление $\mathrm{P} = Q_0 R_1 Q_1 \dots R_n Q_n$, $Q_i \in \{\Sigma \cup \mathcal{V}\}^*$, $R_i \in \mathcal{V}_{\mathrm{e}}{}^{+}\mathcal{V}_{\mathrm{e}}{}^{+}$, такое, что:

- множества переменных образцов Q_i и R_j не пересекаются;
- в каждом слове R_i найдется имеющая единственное вхождение в R_i переменная $\mathtt{e.x_i}$, удовлетворяющая свойству

$$\forall j((R_j = R_{j_1}\, \mathtt{e.x_i}\, \mathtt{R_{j_2}}) \Rightarrow \mathtt{R_j} = \mathtt{R_i}).$$

Этот критерий является необходимым и достаточным условием, если:

- $\mathrm{P} \in \mathcal{V}_{\mathrm{e}}^{+}$;
- или $\mathrm{P} \in \{\mathcal{V}_{\mathrm{e}} \cup \Sigma\}^{+}$ и P линеен.

Пример

Образец $\mathtt{e.x_1}\, \mathtt{e.x_2}\, \mathtt{e.x_1}\, \mathtt{e.x_1}\, \mathtt{e.x_2}\, \mathtt{e.x_1}$ избыточен: $R_i = \mathtt{e.x_1}\, \mathtt{e.x_2}\, \mathtt{e.x_1}$.

Образец $\mathtt{e.x_1}\, \mathtt{e.x_2}\, \mathtt{e.x_2}\, \mathtt{e.x_1}$ является кратким.

Стабильные s-переменные

Определение

Назовем переменную $\mathtt{s.x}$ *в линейном образце* P *стабильной, если*

- $\mathtt{s.x}$ *имеет кратность, не меньшую* 2;
- *или в* P *существует терминальный блок* α *такой, что* $\alpha = \alpha_1\,\mathtt{s.x}\,\alpha_2$, *причем* α_1 *и* α_2 *оба содержат хотя бы один символ или* s*-переменную, имеющую кратность не меньше* 2.

В противном случае назовем $\mathtt{s.x}$ *плавающей.*

Пример

Рассмотрим образец $\mathtt{s.y}_1\ \mathtt{s.y}_2\ \mathtt{e.x}_1\ \mathtt{s.y}_3\ \mathtt{s.y}_4\ \mathtt{s.y}_2\ \mathtt{e.x}_2\ \mathtt{s.y}_5$.
Последовательность его терминальных блоков есть
$\triangleright\,\mathtt{s.y}_1\ \mathtt{s.y}_2 \| \mathtt{s.y}_3\ \mathtt{s.y}_4\ \mathtt{s.y}_2 \| \mathtt{s.y}_5\,\triangleleft$. Стабильными s-переменными являются $\mathtt{s.y}_2$ и $\mathtt{s.y}_1$.

Образцы со стабильными s-переменными

Предложение (Эквивалентность образцов)

Пусть образец P *краткий, линейный, и все его* s*-переменные стабильны. Тогда всякий образец* P′ *такой, что* $\mathscr{L}(\mathrm{P}') = \mathscr{L}(\mathrm{P})$*, либо длиннее* P*, либо получается из* P *переименовывающей подстановкой.*

Построим распознающие подстановки σ_1, σ_2 такие, что $\forall \mathtt{s.x}(\sigma_\mathtt{i}(\mathtt{s.x}) = \xi_{\mathtt{s.x}})$, где $\xi_{\mathtt{s.x}} \in \Sigma$ и ранее не использована (свежая), а $\forall \mathtt{e.x}(\sigma_\mathtt{i}(\mathtt{e.x}) = \xi_{1,\mathtt{e.x}} \dots \xi_{\mathtt{n},\mathtt{e.x}})$, где все $\xi_{i,\mathtt{e.x}} \in \Sigma$ свежие, а n больше, чем длина P. Если $\mathscr{L}(\mathrm{P}') = \mathscr{L}(\mathrm{P})$, то слово $\sigma_1(\mathrm{P}')$ распознается образцом P, слово $\sigma_2(\mathrm{P})$ распознается образцом P′, и последовательности терминальных блоков образцов P и P′ совпадают.

Предложение (Сводимость образцов)

Если образцы P_1 *и* P_2 *линейные и все их* s*-переменные стабильны, то* P_1 *сводится к* P_2 $\Leftrightarrow$ *существует подстановка* σ *такая, что* $\sigma(\mathrm{P}_2) = \mathrm{P}_1$.

Идея доказательства аналогична.

Сложности с плавающими переменными

Если в образце появляются плавающие s-переменных, многие полезные свойства линейных образцов не выполняются!

- Исчезает равносильность сводимости наличию подстановки.
- Исчезает единственность краткой формы.

Известный пример

Образцы e.x s.y и s.y e.x.

Нормальная форма образца

Считаем, что линейный образец P *в нормальной форме*, если

- P не содержит двух и более идущих подряд e-переменных;
- каждая плавающая s-переменная в P предшествует e-переменной и следует за e-переменной.

Всякому линейному образцу P соответствует (с точностью до переименовки переменных) ровно один образец P′ в нормальной форме такой, что $\mathscr{L}(\mathrm{P}) = \mathscr{L}(\mathrm{P}')$.
Образец P′ назовем нормальной формой образца P.

Пример

Нормальная форма образца

$$\mathtt{s.y_1\ s.y_2\ e.x_1\ s.y_3\ s.y_4\ s.y_2\ e.x_2\ s.y_5}$$

есть

$$\mathtt{s.y_1\ s.y_2\ e.x_1\ s.y_3\ e.x_3\ s.y_4\ e.x_4\ s.y_2\ e.x_2\ s.y_5\ e.x_5}.$$

Предложение (Сводимость образцов в нормальной форме)

Если линейные образцы P_1 *и* P_2 *оба в нормальной форме, то* $\mathscr{L}(\mathsf{P}_1) \subseteq \mathscr{L}(\mathsf{P}_2)$ *эквивалентно существованию подстановки* σ *такой, что* $\sigma(\mathsf{P}_2) = \mathsf{P}_1$*.*

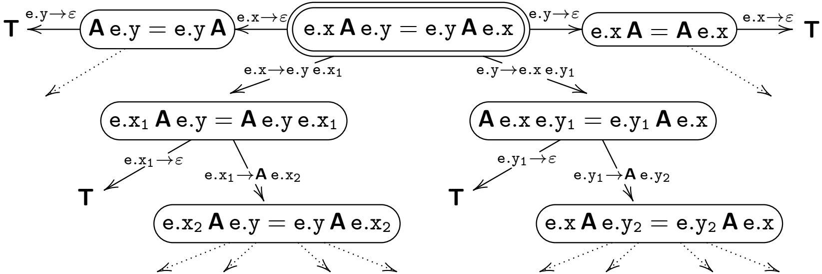
Дерево развертки уравнения в словах

Дано уравнение в словах $\Phi_1^0 = \Phi_2^0$, $\Phi_i \in \{\Sigma \cup \mathcal{V}\}^*$. Помещаем это уравнение в корень дерева.

Правила развертки узла, содержащего уравнение $t_1\,\Phi_1 = t_2\,\Phi_2$:

- Пусть $t_1 = t_2$. Заменяем уравнение в узле на $\Phi_1 = \Phi_2$. Уравнение $\Psi_1 = \Psi_2$, где $\Psi_i \in \{\mathcal{V}_\mathrm{e}\}^*$, $\Psi_j = \varepsilon$, объявляем листом и помечаем **T**.
- Пусть $t_i = \mathtt{s.x}$, $t_j = \mathbf{A}$. Строим потомок $\sigma(\Phi_1) = \sigma(\Phi_2)$, где $\sigma(\mathtt{s.x}) = \mathbf{A}$, для остальных $v \in \mathcal{V}$ $\sigma(v) = v$. Аналогично для $t_j = \mathtt{s.y}$.
- Пусть $t_i = \mathtt{e.x}$, $t_j = \mathbf{A}$. Строим два потомка, соответствующих подстановкам $\sigma_1(\mathtt{e.x}) = \varepsilon$, $\sigma_2(\mathtt{e.x}) = \mathbf{A}\,\mathtt{e.x}_0$, $\mathtt{e.x}_0$ — свежая. Аналогично для $t_j = \mathtt{s.y}$.
- Пусть $t_i = \mathtt{e.x}$, $t_j = \mathtt{e.y}$. Строим потомки по подстановкам $\sigma_1(\mathtt{e.x}) = \mathtt{e.y}\,\mathtt{e.x}_0$, $\sigma_2(\mathtt{e.y}) = \mathtt{e.x}\,\mathtt{e.y}_0$, $\sigma_3(\mathtt{e.x}) = \varepsilon$, $\sigma_4(\mathtt{e.y}) = \varepsilon$.

Особенности дерева развертки уравнения в словах



- В общем случае дерево бесконечно (даже при построении обратного ребра при повторе уравнений на ветви).
- Если каждая e-переменная имеет кратность 1, дерево конечно.

Смысл дерева развертки уравнения в словах

- Дерево развертки уравнения
$$\Phi_1(\mathtt{e.x}_1, \ldots, \mathtt{e.x}_\mathtt{n}) = \Phi_2(\mathtt{e.y}_1, \ldots, \mathtt{e.y}_\mathtt{m})$$
содержит листы **T** $\Leftrightarrow$ верна формула
$$\exists \mathtt{e.x}_1, \ldots, \mathtt{e.x}_\mathtt{n}, \exists \mathtt{e.y}_1, \ldots, \mathtt{e.y}_\mathtt{m} \\ (\Phi_1(\mathtt{e.x}_1, \ldots, \mathtt{e.x}_\mathtt{n}) = \Phi_2(\mathtt{e.y}_1, \ldots, \mathtt{e.y}_\mathtt{m}))$$
- Любое решение уравнения соответствует пути от корня к листу **T**.

Более простое дерево развертки этого уравнения решает задачу поиска отождествляющей подстановки:
$$\exists \mathtt{e.x}_1, \ldots, \mathtt{e.x}_\mathtt{n}, \forall \mathtt{e.y}_1, \ldots, \mathtt{e.y}_\mathtt{m} \\ (\Phi_1(\mathtt{e.x}_1, \ldots, \mathtt{e.x}_\mathtt{n}) = \Phi_2(\mathtt{e.y}_1, \ldots, \mathtt{e.y}_\mathtt{m}))$$

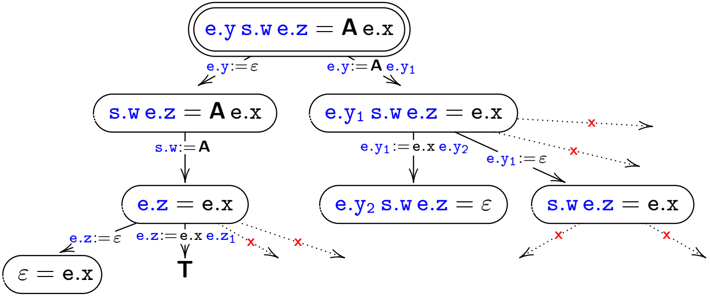
Усеченное дерево развертки уравнения в словах

Дано уравнение в словах $\Phi_1^0 = \Phi_2^0$, у которого множества переменных Φ_1^0 и Φ_2^0 не пересекаются. Множество переменных выражения Φ_1^0 назовем $\mathcal{V}_{\mathrm{subst}}$. Помещаем уравнение $\Phi_1^0 = \Phi_2^0$ в корень дерева.

Правила развертки узла, содержащего уравнение $t_1\,\Phi_1 = t_2\,\Phi_2$:

- Пусть $t_1 = t_2$. Заменяем уравнение в узле дерева на $\Phi_1 = \Phi_2$. Уравнение $\Psi = \varepsilon$, где $\Psi \in \{\mathcal{V}_{\mathsf{e}} \cap \mathcal{V}_{\mathrm{subst}}\}^*$, объявляем листом и помечаем **T**.
- Пусть $t_1 = \mathrm{s.x}$, $t_2 = \mathbf{A}$, $\mathrm{s.x} \in \mathcal{V}_{\mathrm{subst}}$. Строим потомок по подстановке $\sigma(\mathrm{s.x}) = \mathbf{A}$. Аналогично для $t_2 = \mathrm{s.y}$.
- Пусть $t_1 = \mathrm{e.x}$, $t_2 = \mathbf{A}$, $\mathrm{e.x} \in \mathcal{V}_{\mathrm{subst}}$. Строим два потомка, соответствующих подстановкам $\sigma_1(\mathrm{e.x}) = \varepsilon$, $\sigma_2(\mathrm{e.x}) = \mathbf{A}\,\mathrm{e.x}_0$, $\mathrm{e.x}_0$ — свежая, и добавляем $\mathrm{e.x}_0$ в множество $\mathcal{V}_{\mathrm{subst}}$. Аналогично для $t_2 = \mathrm{s.y}$.
- Пусть $t_1 = \mathrm{e.y}$, $t_2 = \mathrm{e.x}$, $\mathrm{e.y} \in \mathcal{V}_{\mathrm{subst}}$. Строим потомки по подстановке $\sigma(\mathrm{e.y}) = \varepsilon$ и по подстановке $\sigma(\mathrm{e.y}) = \mathrm{e.x}\,\mathrm{e.y}_0$, где $\mathrm{e.y}_0$ — свежая, и добавляем $\mathrm{e.y}_0$ в множество $\mathcal{V}_{\mathrm{subst}}$.

Свойства усеченного дерева

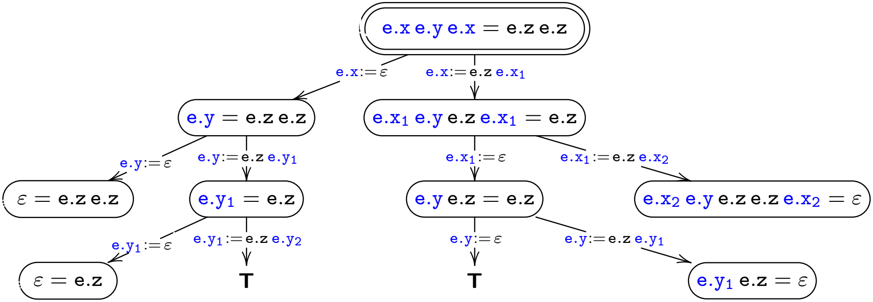


- Уравнение решается относительно переменных из $\mathcal{V}_{\text{subst}}$, остальные переменные рассматриваются как неделимые и не сопоставимые ни с чем, кроме себя, параметры.
- При повторении уравнения на ветви (с точностью до переименовки) разрешим построение обратных ребер.

  Тогда соответствующий граф конечен.

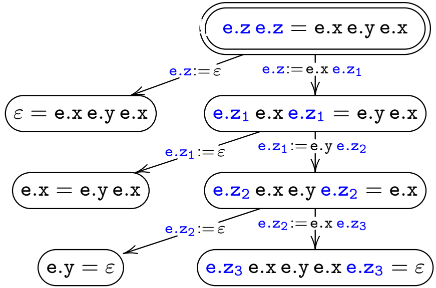
Примеры усеченного дерева

Если решать уравнение $\mathtt{e.x}\,\mathtt{e.y}\,\mathtt{e.x} = \mathsf{e.z}\,\mathsf{e.z}$ по ветвлению переменных $\mathtt{e.x}$ и $\mathtt{e.y}$, мы построим усеченное дерево с двумя листами **T**, соответствующими подстановкам $\{\mathtt{e.x}{:=}\mathsf{e.z}, \mathtt{e.y}{:=}\varepsilon\}$ и $\{\mathtt{e.x}{:=}\varepsilon, \mathtt{e.y}{:=}\mathsf{e.z}\,\mathsf{e.z}\}$.



Примеры усеченного дерева

Если то же самое уравнение e.z e.z = e.x e.y e.x решается по ветвлению переменных e.z, получается другое усеченное дерево развертки, не имеющее ветвей, оканчивающихся значениями **T**.



Действительно, пусть e.x = **A**, e.y = **B**. Нет значения e.z, удовлетворяющего уравнению e.z e.z = **A B A**.

О роли усеченного дерева

Пусть P_1, P_2 — линейные образцы в нормальной форме.

Предложение (Усеченное дерево и подстановочность)

1. *Если существует подстановка σ такая, что $\sigma(\mathsf{P}_1) = \mathsf{P}_2$, то усеченное дерево развертки уравнения $\mathsf{P}_1 = \mathsf{P}_2$ содержит листы* **T**.
2. *Верно и обратное.*

Предложение (Усеченное дерево и сводимость)

P_2 сводится к P_1 тогда и только тогда, когда в усеченном дереве развертки уравнения $\mathsf{P}_1 = \mathsf{P}_2$ есть листы **T**.

А с повторными e-переменными?

Предложение

Пусть $\mathsf{P}_1, \mathsf{P}_2 \in \{\Sigma \cup \mathcal{V}_\mathsf{e}\}^+$.

1. *Подстановка* σ *такая, что* $\sigma(\mathsf{P}_1) = \mathsf{P}_2$, *существует тогда и только тогда, когда усеченное дерево развертки уравнения* $\mathsf{P}_1 = \mathsf{P}_2$ *содержит листы* $\mathbf{T}$.
2. *Образец* P_1 *сводится к образцу* P_2 *тогда и только тогда, когда существует подстановка* σ *такая, что* $\sigma(\mathsf{P}_2) = \mathsf{P}_1$.

Итак:

- Свойство подстановочности выполняется для многих классов плоских образцов Рефала.
- Полезно для удаления избыточных предложений и при учете отрицательных условий в суперкомпиляции.

Ограничения

- Всюду в докладе алфавит Σ бесконечен в смысле наличия достаточного количества букв, не входящих в образцы P_1 и P_2. Для $|\Sigma| \notin \{1, \infty\}$ проблема сводимости неразрешима! (Freidenberger, Reidenbach, 2014)
- Дизъюнкция и конъюнкция двух образцов в общем случае не представимы в языке образцов.