Обобщение при уточнении понятия замкнутой переменной

А. Н. Непейвода
ИПС им. А.К. Айламазяна РАН

V совместное совещание по языку Рефал

16 июня 2022, МГТУ им. Н.Э. Баумана

Наводящий пример

При обобщении $(\varepsilon)(e.x)$ и $(A)(A\ e.x')$ получается заготовка $(e.x_1)(e.x_1\ e.x_2)$. После чего параметры $e.x_1$ и $e.x_2$ сливаются как подряд идущие, даже если на $e.x$ и $e.x'$ есть рестрикция, запрещающая вхождения буквы A, переносимая на $e.x_2$.

Частичное решение: записать уравнение, устанавливающее связь между переменными после слияния — оказывается слабым, потому что негативные рестрикции всё равно теряются.

Идея

Перенести часть зависимостей из языка рестрикций (уравнения) в язык конфигурации (параметризованное выражение).

Что сейчас:

- Строится кандидат общего вида без обобщений вида $e.i = \varepsilon \mid \varepsilon$.
- Сливаются подряд идущие e-параметры.
- Выбирается наилучший по оценке.

Что хочется:

- Строится кандидат общего вида.
- Если он безопасный $\Rightarrow$ OK, иначе параметры сливаются, чтобы стал безопасным .
- Выбирается наилучший по оценке.

Идея

Перенести часть зависимостей из языка рестрикций (уравнения) в язык конфигурации (параметризованное выражение).

Что было:

- Строится кандидат общего вида без обобщений вида $e.i = \varepsilon \mid \varepsilon$.
- Сливаются подряд идущие e-параметры.
- Выбирается наилучший по оценке.

Как стало:

- Строится кандидат общего вида.
- Безопасный $\Rightarrow$ OK, иначе параметры сливаются, чтобы стал безопасным (future work).
- Выбирается наилучший по оценке.

Якорные фрагменты

(пока что в допущении, что t-параметров нет, и без учёта общезначимости дизъюнкции: $e.i = s.i\, e.j \lor e.i = (e.j_1)e.j_2$)

- Константный фрагмент — дерево, не содержащее e-параметров.
- Плоское разбиение — n-ка подвыражений, помечающих узлы в лесе структуры (структурные скобки + вызовы функций) без учёта константных фрагментов.
- Свободный фрагмент — максимальное подвыражение, содержащее только e-параметры.
- Якорный фрагмент — подвыражение плоского разбиения, стоящее между e-параметрами и не содержащее e-параметров.

Язык образца

Базовое семантическое понятие для построения обобщения ассоциативных данных.

Пусть $\mathscr{P}$ — образец; тогда языком $\mathscr{L}(\mathscr{P})$ называется множество всех возможных слов, которые получаются некоторой подстановкой в $\mathscr{P}$.

Если $\mathscr{P}'$ — обобщение $\mathscr{P}$, и $\mathscr{P}$ и $\mathscr{P}'$ имеют одну и ту же структуру леса над плоским разбиением, то $\mathscr{L}(\mathscr{P}) \subseteq \mathscr{L}(\mathscr{P}')$ (как образцов, полученных из параметризованных выражений).

Возможные проблемы

Что плохое может случиться, если мы перестанем сливать соседние параметры?

Перестаёт выполняться критерий обрыва цепи вычислений (нарушается свойство wqo).

-
    - Крускал + Турчин $\Rightarrow$ безопасно (Крускал выполняется всегда, а Турчин не смотрит на параметры).

Возможные проблемы

Что плохое может случиться, если мы перестанем сливать соседние параметры?

Перестаёт выполняться критерий обрыва цепи вычислений (нарушается свойство wqo).

- Крускал + Турчин $\Rightarrow$ безопасно (Крускал выполняется всегда, а Турчин не смотрит на параметры).

Последовательность обобщений становится бесконечной (нарушается свойство нётеровости).

- Опасно!

$$P_1 = e.x_0 e.x_0$$
$$P_2 = e.x_1 e.x_0 e.x_0 e.x_1$$
$$P_3 = e.x_2 e.x_1 e.x_0 e.x_0 e.x_1 e.x_2$$

Открытые переменные

Рассмотрим произвольный образец $\mathtt{P}$.

- Семантически открытая $\mathtt{e}$-переменная — такая переменная $\mathtt{e.i}$, что при сопоставлении вида $\mathtt{P,\ e.i} : \mathtt{Cond}$ могут потребоваться рекурсивные возвраты (расширения).
- Синтаксически открытая $\mathtt{e}$-переменная — переменная, входящая в некоторый элемент плоского разбиения $\mathtt{P}$ вместе с какой-нибудь другой $\mathtt{e}$-переменной.

Совпадают ли эти понятия?

Синтаксис vs семантика

Рассмотрим образец $(e.x_1\ e.x_2)\ e.x_2\ e.x_1\ e.x_1$. Он однозначный (т.е. существует единственная возможная подстановка), но в нём нет ни одной синтаксически замкнутой переменной.

- Проводим разбиение на равносоставленные фрагменты, отчего длина $e.x_1$ определяется однозначно.

Рассмотрим образец $(e.x_1\ e.x_2)(e.x_2\ e.x_3)(e.x_3\ e.x_1)$. Он однозначный, но в элементах его плоского разбиения нет равносоставленных префиксов или суффиксов.

- Сравнение длин двух решений приводит к противоречию.

$\mathfrak{n}$-замкнутость

Пусть $\mathscr{P}$ — образец (выражение); $P_1, \ldots, P_n$ — свободные фрагменты элементов его плоского разбиения (далее кратко СФР).

- Если $e.j$ — единственная e-переменная, входящая в некоторый P_k, тогда $e.j$ — 0-замкнутая.
- Если $e.j_1, \ldots, e.j_k$ — j_i-замкнутые переменные, входящие в P_k вместе с некоторой $e.j_{k+1}$, степень замкнутости которой неизвестна либо больше $\max(j_i) + 1$, тогда $e.j_{k+1}$ — $\max(j_i) + 1$-замкнутая.

Утверждение

Если все переменные образца $\mathscr{P}$ i-замкнутые, тогда степень замкнутости переменной, входящей в его СФР $P_1, \ldots, P_n$, не может превышать $n - 1$.

Мера открытости

Дан СФР P. n-ку $(t_{n-1}, \ldots, t_0)$ такую, что t_i — количество различных переменных замкнутости i, входящих в P, назовём мерой открытости $\mu(P)$.

В образце $(\mathsf{e.1}\ \mathsf{e.1})(\mathsf{e.2})(\mathsf{e.1}\ \mathsf{e.2})$ все переменные являются 0-замкнутыми (вычисляются однозначно), но меры открытости его СФР $(P_1,\ P_2,\ P_3) = (\mathsf{e.1}\ \mathsf{e.1},\ \mathsf{e.2},\ \mathsf{e.1}\ \mathsf{e.2})$ различны. $\mu(P_1) = \mu(P_2) = (0, 0, 1)$, $\mu(P_3) = (0, 0, 2)$

Мера открытости

Дан СФР P. n-ку $(t_{n-1}, \ldots, t_0)$ такую, что t_i — количество различных переменных замкнутости i, входящих в P, назовём мерой открытости $\mu(P)$.

В образце $(e.1\, e.1)(e.2)(e.1\, e.2)$ все переменные являются 0-замкнутыми (вычисляются однозначно), но меры открытости его СФР $(P_1,\ P_2,\ P_3) = (e.1\, e.1,\ e.2,\ e.1\, e.2)$ различны. $\mu(P_1) = \mu(P_2) = (0, 0, 1)$, $\mu(P_3) = (0, 0, 2)$

Утверждение

Если в СФР i-замкнутого образца добавляется e-переменная, которая туда не входит (но может входить в другие СФР того же образца), то мера открытости этого СФР может только увеличиться лексикографически.

Различающая подстановка

Пусть Φ_i — константные фрагменты элементов плоского образца P. Рассмотрим различные $\Psi_i \in (\Sigma^*)$ такие, что $\forall i, j(\Psi_i \neq \Phi_j)$ (таких Ψ_i можно построить неограниченное количество). Пусть $e.i$ — e-переменные, входящие в P. Назовём подстановку σ такую, что $\sigma(e.i) = \Psi_i$, различающей подстановкой.

Утверждение

Пусть $(P_1, \ldots, P_n)$ — кортеж СФР k-замкнутого образца $\mathscr{P}$. Построим образец $\mathscr{P}'$ добавлением в некоторые P_i ещё одного вхождения какой-либо переменной, уже входящей в P_i. Тогда $\mathscr{L}(\mathscr{P}) \nsubseteq \mathscr{L}(\mathscr{P}')$.

Оставшийся случай

Набор переменных в P_i остался тот же, при этом кратность некоторых из них увеличилась, а других — уменьшилась. Может ли такое преобразование (как минимум) сохранять язык образца?

- По индукции: для СФР, имеющих $\mu(P) = (0, \dots, 0, 1)$, такое невозможно из-за существования различающей подстановки. Следовательно, невозможно для P с $\mu(P) = (0, \dots, 0, n)$.
- Предположим, что для всех СФР, имеющих $\mu(P) = (0, \dots, 0, t_i, \dots)$, такое невозможно. Тогда для СФР, имеющих $\mu(P) = (0, \dots, 1\ t_i, \dots)$, это также невозможно, поскольку значения всех прочих переменных после различающей подстановки фиксированы.

Нётеровость обобщений

Набросок теоремы

Если k_i-замкнутый образец обобщается до k_j-замкнутого, тогда выполняется минимум одно условие (в порядке приоритетов):

- уменьшается число элементов плоского разбиения или СФР;
- увеличивается общее число различных переменных (ограниченное сверху числом СФР);
- увеличивается мера открытости СФР (также ограничена сверху);
- при сохранении двух предыдущих мер — уменьшается кратность некоторой переменной.

Future Work

- Точный критерий семантической замкнутости переменной либо доказательство алгоритмической неразрешимости задачи проверки семантической замкнутости (и поиск различных эвристик, аппроксимирующих её снизу).
- Модификация алгоритмов сопоставления с образцом с учётом расширенного понятия замкнутости.
- Совершенствование алгоритма приведения обобщённого выражения к «безопасной» форме.

Future Work

- Точный критерий семантической замкнутости переменной либо доказательство алгоритмической неразрешимости задачи проверки семантической замкнутости (и поиск различных эвристик, аппроксимирующих её снизу).
- Модификация алгоритмов сопоставления с образцом с учётом расширенного понятия замкнутости.
- Совершенствование алгоритма приведения обобщённого выражения к «безопасной» форме.

Спасибо за внимание!