Совместный рабочий семинар
МГТУ им. Н.Э. Баумана и
ИПС им. А.К. Айламазяна РАН, 19 декабря, 2025

О некоторых классах плоских образцов с повторными переменными

А. Непейвода, a_nevod@mail.ru

Языки образцов и проблема слова

В данном докладе под переменными всегда понимаются e-переменные в смысле Рефала.

Плоский образец — строка из терминальных символов (Σ) и переменных (X): $\alpha = x_1 \mathbf{a} x_2 \mathbf{b} x_1$ (пример)
Язык образца (Pattern Language) — множество всех слов, полученных подстановкой строк вместо переменных:

$$L(\alpha) = \{h(\alpha) \mid h : X \to \Sigma^* \text{ — подстановка}\}$$

Дано: образец α, слово $w \in \Sigma^*$.
Вопрос: $w \in L(\alpha)$?

В общем случае для образцов *NP-полна* (Angluin, 1980)

Идея доказательства теоремы Англуин

Конструкция, сводящая к проблеме 1-in-3-SAT (проблеме выполнимости КНФ с тремя литералами в дизъюнкте, ровно один из которых истинен):

- нужно промоделировать дизъюнкты строками, литералы — их подстроками, при этом литералы должны быть перестановочными (коммутативность дизъюнкции)...
- если $uw = wu$, то что можно сказать про значения u и w?
- нужно вынудить строки принимать только два различных значения, при этом если некоторая переменная принимает значение u, то «сопряжённая» переменная обязана принимать значение w.
- нужно вынудить дизъюнкт принимать значение «не меньше» одной единицы.

Идея доказательства теоремы Англуин

- Каждой переменной V_i сопоставим переменную образца x_i и двойственную ей y_i.
- Каждому положительному литералу с переменной V_j сопоставим p_j, отрицательному — q_j.
- Дизъюнкты разделим константами $\mathbf{b}$.
- Факт двойственности выразим условием: $x_i y_i$ сопоставляется с $\mathbf{a}$. Тогда одна из них обязана принять значение ε (FALSE), а вторая — $\mathbf{a}$ (TRUE).
- Факт 1-in-3-SAT выразим условием: образ дизъюнкта $L_1 \vee L_2 \vee L_3$ сопоставляется с $\mathbf{a}$. Тогда образ ровно одного литерала принимает значение $\mathbf{a}$, остальные — ε.

Искомый образец:

$$\mathbf{b}\,x_1 y_1 \mathbf{b}\,x_2 y_2 \dots \mathbf{b}\,x_k y_k$$
$$\mathbf{b}(L_{1,1}\sigma)(L_{1,2}\sigma)(L_{1,3}\sigma)\mathbf{b} \dots \mathbf{b}(L_{n,1}\sigma)(L_{n,2}\sigma)(L_{n,3}\sigma)$$

Искомая строка: $(\mathbf{ba})^{k+n}$

Идея доказательства теоремы Англуин

Искомый образец:

$$\mathbf{b}\, x_1 y_1 \mathbf{b}\, x_2 y_2 \ldots \mathbf{b}\, x_k y_k$$
$$\mathbf{b}(L_{1,1}\sigma)(L_{1,2}\sigma)(L_{1,3}\sigma)\mathbf{b} \ldots \mathbf{b}(L_{n,1}\sigma)(L_{n,2}\sigma)(L_{n,3}\sigma)$$

Искомая строка: $(\mathbf{ba})^{k+n}$

Например, для 1-in-3-SAT КНФ

$$(P \vee Q \vee R) \;\&\; (P \vee \neg Q \vee \neg R) \;\&\; (\neg P \vee Q \vee \neg R)$$

получается задача сопоставления образца:

$$\mathbf{b}\, x_P y_P \mathbf{b}\, x_Q y_Q \mathbf{b}\, x_R y_R \mathbf{b}\, x_P x_Q x_R \mathbf{b}\, x_P y_Q y_R \mathbf{b}\, y_P x_Q y_R$$

со строкой $\mathbf{b\,a\,b\,a\,b\,a\,b\,a\,b\,a\,b\,a}$.

Наблюдение

Образцы в доказательстве очень спутанные и нагружены большим количеством повторных вхождений.

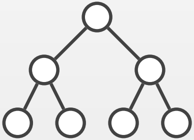
Существуют ли более практичные классы плоских образцов, допускающие полиномиальное сопоставление и в реальности чаще используемые в Рефале?

Частичный ответ: статья Reidenbach, D., & Schmid, M.L.: *Patterns with bounded treewidth*, **Information and Computation, 2014**

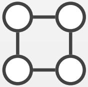
Что такое древесная ширина?

Мера того, насколько граф похож на дерево:

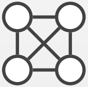
- Деревья имеют древесную ширину (treewidth) 1
- Циклические графы имеют tw = 2
- Полные графы K_n имеют tw = $n - 1$



Treewidth = 1



Treewidth = 2



Treewidth = 3

Формальное определение древесной ширины

Древовидное разложение (Tree Decomposition)

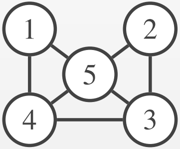
Для графа $G = (V, E)$ — пара $(T, \{B_t\}_{t \in T})$, где T — дерево, $B_t \subseteq V$ — мешки (bags), такие что:

1. $\bigcup_{t \in T} B_t = V$
2. $\forall \{u, v\} \in E \; \exists t \in T : u, v \in B_t$
3. $\forall v \in V$ множество $\{t \in T \mid v \in B_t\}$ связно в T

- Ширина разложения: $\max_{t \in T} |B_t| - 1$
- *Древесная ширина графа* $tw(G)$ — минимальная ширина среди всех древовидных разложений G

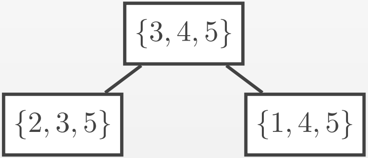
Формальное определение древесной ширины

- Ширина разложения: $\max_{t\in T} |B_t| - 1$
- *Древесная ширина графа* $tw(G)$ — минимальная ширина среди всех древовидных разложений G



Исходный граф

$\Rightarrow$



Древовидное разложение (width = 2)

Ограниченная ширина = эффективные алгоритмы

Если $tw(G) \leq k$, то многие NP-трудные задачи на G решаются за полиномиальное время:

- Задача о гамильтоновом пути
- Максимальная клика

Динамическое программирование на древовидном разложении:

- Обходим дерево разложения снизу вверх
- В каждом мешке храним частичные решения для вершин в этом мешке
- Так как размер мешка $\leq k + 1$, состояний — $O(n^{k+1})$
- Слияние решений вдоль дерева — полиномиальное время

Кодирование образца как системы отношений

- E — бинарное: соединяет вхождения одной переменной
- S — бинарное: перечисление всех смежных позиций
- L, R, D_b — унарные: левый/правый конец, позиции с терминалом b

Структура $\mathcal{A}_\alpha$ для образца α

- Универсум: $\{1, 2, \ldots, |\alpha|\}$ (позиции в α)
- $S^{\mathcal{A}_\alpha} = \{(i, i+1) \mid 1 \leq i < |\alpha|\}$
- $E^{\mathcal{A}_\alpha}$: минимальное отношение, чьё симметрично-транзитивное замыкание соединяет все вхождения каждой переменной
- $L^{\mathcal{A}_\alpha} = \{1\}, R^{\mathcal{A}_\alpha} = \{|\alpha|\}$
- $D_b^{\mathcal{A}_\alpha} = \{i \mid \alpha[i] = b\}$

$\alpha = x_1\mathbf{a}x_1\mathbf{b}x_2$: позиции 1 и 3 соединены через E, позиции 2 и 4 — терминалы $\mathbf{a}$ и $\mathbf{b}$, S соединяет (1,2), (2,3), (3,4), (4,5)

Кодирование слова

Структура $\mathcal{A}_w$ для слова w (случай без пустых подстановок):

- Универсум: $\{(i,j) \mid 1 \leq i \leq j \leq |w|\}$ (все подстроки)
- $S^{\mathcal{A}_w} = \{((i,j),(j+1,k))\}$ (конкатенация подстрок)
- $E^{\mathcal{A}_w} = \{((i,j),(i',j')) \mid w[i..j] = w[i'..j']\}$ (равные подстроки)
- $L^{\mathcal{A}_w}$ — все префиксы, $R^{\mathcal{A}_w}$ — все суффиксы
- $D_b^{\mathcal{A}_w} = \{(i,i) \mid w[i] = b\}$

Лемма о редукции

Лемма

Для любого образца α и слова w:

$$w \in L(\alpha) \Leftrightarrow \exists \text{ гомоморфизм } h : \mathcal{A}_\alpha \to \mathcal{A}_w$$

Гомоморфизм h отображает позиции α в подстроки w, сохраняя:

- Смежность $(S) \Rightarrow$ конкатенацию подстрок
- Равенство переменных $(E) \Rightarrow$ равенство соответствующих подстрок
- Терминалы $(D_b) \Rightarrow$ символы b
- Границы $(L, R) \Rightarrow$ покрытие всего слова w

Граф Гафмана (Gaifman graph) структуры

Для структуры $\mathcal{A}$ граф $G(\mathcal{A})$:

- Вершины — элементы универсума
- Ребро между u и v, если $\exists R \in \tau$ и кортеж в $R^{\mathcal{A}}$, содержащий u и v

$tw(\mathcal{A}) := tw(G(\mathcal{A}))$

Связь tw и проблемы слова

Пусть P — класс образцов, для которых существует полиномиально вычислимое кодирование в структуры $\mathcal{A}_\alpha$ с ограниченной treewidth:

$$\exists k \, \forall \alpha \in P : tw(\mathcal{A}_\alpha) \leq k$$

Тогда проблема слова для него разрешима за полиномиальное время.

Доказательство:

1. По Лемме о редукции: $w \in L(\alpha) \Leftrightarrow \exists$ гомоморфизм $\mathcal{A}_\alpha \to \mathcal{A}_w$
2. По теореме Фройдера: для структур с ограниченной treewidth проверка существования гомоморфизма — полиномиальна

Умеренно спутанные образцы

- *Спутанные (entwined)* переменные: x и y спутаны в образце, если встречаются в его подслове вида $x\gamma_1 y\gamma_2 x\gamma_3 y$, $\gamma_i \in (X \cup \Sigma)^*$.
- *Тесно спутанный (closely entwined)*: если x и y спутаны, причём $|\gamma_2|_x = |\gamma_2|_y = 0$, тогда $\gamma_2 = \varepsilon$.
- *Умеренно спутанный (mildly entwined)*:
  1. тесно спутанный,
  2. фрагмент между двумя последовательными вхождениями любой переменной — вложенный (nested) образец.

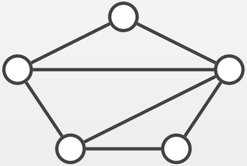
Умеренно спутанный: $x_1\mathbf{a}x_2x_1\mathbf{b}x_2$ (переплетены, но тесно)

Не умеренно спутанный: $x_1x_2x_3x_2x_3x_1$ (между x_1 и x_1 не вложенный фрагмент)

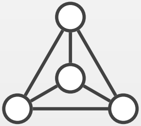
Внешнепланарные графы и их свойства

Граф называется *внешнепланарным* (*outerplanar*), если его можно нарисовать на плоскости так, что:

1. Все вершины лежат на внешней грани
2. Рёбра не пересекаются

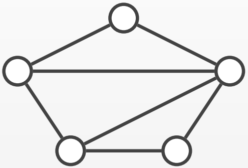


Внешнепланарный ($tw \leq 2$)

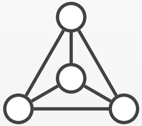


K_4 (Планарный, $tw = 3$)

Внешнепланарные графы и их свойства



Внешнепланарный ($tw \leq 2$)



K_4 (Планарный, $tw = 3$)

Внешнепланарные графы имеют $tw \leq 2$

Связь с умеренно спутанными образцами

Лемма 20, RS

Для любого образца α:

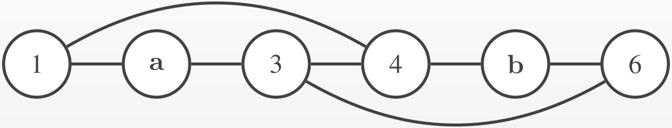
$$\mathcal{G}(\mathcal{A}^s_\alpha) \text{ — внешнепланарный} \Leftrightarrow \alpha \text{ — умеренно спутанный}$$

где $\mathcal{G}(\mathcal{A}^s_\alpha)$ — граф Гафмана стандартной структуры α

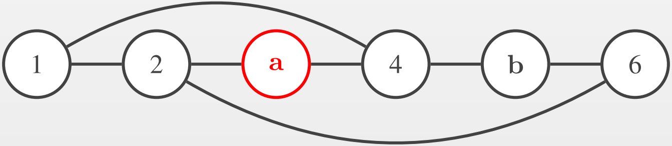
Класс умеренно спутанных образцов имеет $tw \leq 2$, следовательно, проблема слова для него разрешима за полиномиальное время.

Примеры

- Граф Гафмана стандартной структуры умеренно спутанного образца $x_1\mathbf{a}x_2x_1\mathbf{b}x_2$ внешнепланарен:



- Граф Гафмана стандартной структуры сложно спутанного* образца $x_1x_2\mathbf{a}x_1\mathbf{b}x_2$ — появляется внутренняя вершина $\mathbf{a}$:



(* — *Хотя граф образца и не внешнепланарен, его* $tw = 2$.)

Наблюдение-2

- Использование повторных переменных в образцах Рефала (в моих исходных текстах как минимум) — в рамках умеренной спутанности
- Алгоритм построения морфизма Шмида-Рейденбаха расширяется на повторные переменные типа символ (но возможно, есть и более удачные идеи алгоритмов сопоставления, основанные на ДП)

Ещё один класс образцов и наблюдение-3

Образцы с однозначным сопоставлением:

- $x\mathbf{a}yx\mathbf{b}y$ (пример Ливчака)
- Из новой серии: $x^{k+2}(\mathbf{c}_i y)^{m+2}$, $\mathbf{c}_i$ различны
- Также: $xx\mathbf{a}yx\mathbf{b}y$, $xxy\mathbf{a}y\mathbf{b}x$.

Все найденные нами и классические минимальные однозначные образцы внешнепланарны.

Очевидно, можно построить однозначные образцы с нарушением этого свойства (например, $x\mathbf{a}y\mathbf{a}\mathbf{b}x\mathbf{b}y$), однако в них можно стереть некоторые константы и снова получить однозначный образец, т.е. они не минимальные.