Последовательное замыкание классов языков относительно пересечения и гомоморфизмов

Тихон Григорьевич Пшеницын

МГУ, механико-математический факультет

Конференция “Ломоносов”, 11 ноября 2020

Научный руководитель Мати Рейнович Пентус

Суть задачи и откуда она возникла

- Общий вопрос теории формальных грамматик: какой класс языков задают данные грамматики? Особый интерес для нас представляют категориальные грамматики.

Суть задачи и откуда она возникла

- Общий вопрос теории формальных грамматик: какой класс языков задают данные грамматики? Особый интерес для нас представляют категориальные грамматики.
- Пример: категориальные грамматики Ламбека задают класс контекстно-свободных языков без пустого слова.

Суть задачи и откуда она возникла

- Общий вопрос теории формальных грамматик: какой класс языков задают данные грамматики? Особый интерес для нас представляют категориальные грамматики.
- Пример: категориальные грамматики Ламбека задают класс контекстно-свободных языков без пустого слова.
- Классы языков, порождаемых категориальными грамматиками, замкнуты относительно замен символов алфавита на другие (символьных гомоморфизмов, strictly alphabetic morphisms).

Суть задачи и откуда она возникла

- Общий вопрос теории формальных грамматик: какой класс языков задают данные грамматики? Особый интерес для нас представляют категориальные грамматики.
- Пример: категориальные грамматики Ламбека задают класс контекстно-свободных языков без пустого слова.
- Классы языков, порождаемых категориальными грамматиками, замкнуты относительно замен символов алфавита на другие (символьных гомоморфизмов, strictly alphabetic morphisms).
- Существуют расширения исчисления Ламбека с помощью конъюнкции. Можно показать, что языки соответствующих грамматик замкнуты относительно пересечения (Канадзава 1992). Точного описания класса нет.

Суть задачи и откуда она возникла

- Общий вопрос теории формальных грамматик: какой класс языков задают данные грамматики? Особый интерес для нас представляют категориальные грамматики.
- Пример: категориальные грамматики Ламбека задают класс контекстно-свободных языков без пустого слова.
- Классы языков, порождаемых категориальными грамматиками, замкнуты относительно замен символов алфавита на другие (символьных гомоморфизмов, strictly alphabetic morphisms).
- Существуют расширения исчисления Ламбека с помощью конъюнкции. Можно показать, что языки соответствующих грамматик замкнуты относительно пересечения (Канадзава 1992). Точного описания класса нет.
- Вопрос: если последовательно замыкать класс контекстно-свободных языков относительно пересечений, а потом гомоморфизмов, когда мы перестанем получать новые языки?

О языках и алфавитах

Хочется ответить на поставленный вопрос в максимальной общности.

О языках и алфавитах

Хочется ответить на поставленный вопрос в максимальной общности.

Пусть мы работаем внутри счётного алфавита Σ. $\Sigma^* = \bigcup\limits_{n=0}^{\infty} \Sigma^n$.

Определение

Язык над алфавитом Σ — это произвольное подмножество Σ^*.

О языках и алфавитах

Хочется ответить на поставленный вопрос в максимальной общности.

Пусть мы работаем внутри счётного алфавита Σ. $\Sigma^* = \bigcup\limits_{n=0}^{\infty} \Sigma^n$.

Определение

Язык над алфавитом Σ — это произвольное подмножество Σ^*.

Определение

Набор языков $\mathcal{L}$ над алфавитом Σ имеет локально конечный алфавит, если для любого $L \in \mathcal{L}$ множество символов, встречающихся в словах из L, конечно.

Символьные гомоморфизмы и соответствия

Определение

Класс языков $\mathcal{L}$ замкнут относительно ***символьных гомоморфизмов****, если для любого языка $L \in \mathcal{L}$ и для любой функции $f : \Sigma \to \Sigma$ верно, что $f(L)$ принадлежит $\mathcal{L}$.*
Здесь функция f расширяется на слова по правилу $f(a_1 \dots a_n) := f(a_1) \dots f(a_n)$.

Символьные гомоморфизмы и соответствия

Определение

Класс языков $\mathcal{L}$ замкнут относительно ***символьных гомоморфизмов****, если для любого языка $L \in \mathcal{L}$ и для любой функции $f : \Sigma \to \Sigma$ верно, что $f(L)$ принадлежит $\mathcal{L}$.*
Здесь функция f расширяется на слова по правилу $f(a_1 \dots a_n) := f(a_1) \dots f(a_n)$.

Определение

Класс языков $\mathcal{L}$ замкнут относительно ***локально конечных символьных соответствий****, если для любого языка $L \in \mathcal{L}$ и для любого бинарного отношения $R \subseteq \Sigma \times \Sigma$, такого, что $|\{b \mid aRb\}| < \infty$ для всех $a \in \Sigma$, верно, что $R(L)$ принадлежит $\mathcal{L}$, где $R(L) = \{b_1 \dots b_n \mid \exists a_1 \dots a_n \in L : a_i R b_i\}$.*

Замыкание относительно пересечений, гомоморфизмов, соответствий

Замыкание класса языков $\mathcal{L}$ относительно пересечения (символьных гомоморфизмов/локально конечных символьных соответствий) обозначается $[\mathcal{L}]_\cap$ ($[\mathcal{L}]_{\mathrm{Hom}}/[\mathcal{L}]_{\mathrm{LFRel}}$).

Основная теорема

Теорема

Пусть $\mathcal{L}$ — произвольный класс языков с локально конечным алфавитом. Тогда $[[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}}$ замкнут относительно пересечений и локально конечных символьных соответствий.

Основная теорема

Теорема

Пусть $\mathcal{L}$ — произвольный класс языков с локально конечным алфавитом. Тогда $[[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}}$ замкнут относительно пересечений и локально конечных символьных соответствий.

Замечание

Из теоремы следует, что $[[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}} = [[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{LFRel}}$.

Наблюдения для доказательства

Теорема

Пусть $\mathcal{L}$ — произвольный класс языков с локально конечным алфавитом. Тогда $[[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}}$ замкнут относительно пересечений.

- Надо проверить, что пересечение любых двух языков класса $[[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}}$ также принадлежит этому классу.

Наблюдения для доказательства

Теорема

Пусть $\mathcal{L}$ — произвольный класс языков с локально конечным алфавитом. Тогда $[[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}}$ замкнут относительно пересечений.

- Надо проверить, что пересечение любых двух языков класса $[[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}}$ также принадлежит этому классу.
- У семейства $[\mathcal{L}]_{\mathrm{LFRel}}$ локально конечный алфавит.

Наблюдения для доказательства

Теорема

Пусть $\mathcal{L}$ — произвольный класс языков с локально конечным алфавитом. Тогда $[[[\mathcal{L}]_{\mathrm{LFRel}}]_\cap]_{\mathrm{Hom}}$ замкнут относительно пересечений.

- Надо проверить, что пересечение любых двух языков класса $[[[\mathcal{L}]_{\mathrm{LFRel}}]_\cap]_{\mathrm{Hom}}$ также принадлежит этому классу.
- У семейства $[\mathcal{L}]_{\mathrm{LFRel}}$ локально конечный алфавит.
- Произвольный язык из $[[\mathcal{L}]_{\mathrm{LFRel}}]_\cap$ имеет вид $L_1 \cap \dots \cap L_n$, где $L_i \in [\mathcal{L}]_{\mathrm{LFRel}}$.

Иллюстрация доказательства

- Пусть даны четыре языка $L_i = L_i(a, b, c)$, $i = 1, 2, 3, 4$ из класса $[\mathcal{L}]_{\mathrm{LFRel}}$, слова в которых состоят из символов a, b, c.

Иллюстрация доказательства

- Пусть даны четыре языка $L_i = L_i(a, b, c)$, $i = 1, 2, 3, 4$ из класса $[\mathcal{L}]_{\mathrm{LFRel}}$, слова в которых состоят из символов a, b, c.
- Пусть $f_1, f_2 : \Sigma \to \Sigma$ задают символьные гомоморфизмы, причём
  1. $f_1(a) = d$, $f_1(b) = f_1(c) = e$;
  2. $f_2(a) = f_2(b) = e$, $f_2(c) = d$.

Иллюстрация доказательства

- Пусть даны четыре языка $L_i = L_i(a, b, c)$, $i = 1, 2, 3, 4$ из класса $[\mathcal{L}]_{\mathrm{LFRel}}$, слова в которых состоят из символов a, b, c.
- Пусть $f_1, f_2 : \Sigma \to \Sigma$ задают символьные гомоморфизмы, причём
  1. $f_1(a) = d$, $f_1(b) = f_1(c) = e$;
  2. $f_2(a) = f_2(b) = e$, $f_2(c) = d$.
- Хотим: $f_1(L_1 \cap L_2) \cap f_2(L_3 \cap L_4) \in [[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}}$.

Иллюстрация доказательства

- Пусть даны четыре языка $L_i = L_i(a, b, c)$, $i = 1, 2, 3, 4$ из класса $[\mathcal{L}]_{\mathrm{LFRel}}$, слова в которых состоят из символов a, b, c.
- Пусть $f_1, f_2 : \Sigma \to \Sigma$ задают символьные гомоморфизмы, причём
  1. $f_1(a) = d$, $f_1(b) = f_1(c) = e$;
  2. $f_2(a) = f_2(b) = e$, $f_2(c) = d$.
- Хотим: $f_1(L_1 \cap L_2) \cap f_2(L_3 \cap L_4) \in [[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}}$.
- $|f_1^{-1}(e)| = |f_2^{-1}(e)| = 2$, $|f_1^{-1}(d)| = |f_2^{-1}(d)| = 1$.

Иллюстрация доказательства

- Пусть даны четыре языка $L_i = L_i(a, b, c)$, $i = 1, 2, 3, 4$ из класса $[\mathcal{L}]_{\mathrm{LFRel}}$, слова в которых состоят из символов a, b, c.
- Пусть $f_1, f_2 : \Sigma \to \Sigma$ задают символьные гомоморфизмы, причём
  1. $f_1(a) = d$, $f_1(b) = f_1(c) = e$;
  2. $f_2(a) = f_2(b) = e$, $f_2(c) = d$.
- Хотим: $f_1(L_1 \cap L_2) \cap f_2(L_3 \cap L_4) \in [[[\mathcal{L}]_{\mathrm{LFRel}}]_{\cap}]_{\mathrm{Hom}}$.
- $|f_1^{-1}(e)| = |f_2^{-1}(e)| = 2$, $|f_1^{-1}(d)| = |f_2^{-1}(d)| = 1$.
- Выделяем в Σ $2 \times 2 = 4$ символа: e_{ba}, e_{bb}, e_{ca}, e_{cb}. Для порядка добавляем еще символ d_{ac}.

Иллюстрация доказательства

- Пусть даны четыре языка $L_i = L_i(a, b, c)$, $i = 1, 2, 3, 4$ из класса $[\mathcal{L}]_{\mathrm{LFRel}}$, слова в которых состоят из символов a, b, c.
- Пусть $f_1, f_2 : \Sigma \to \Sigma$ задают символьные гомоморфизмы, причём
  1. $f_1(a) = d$, $f_1(b) = f_1(c) = e$;
  2. $f_2(a) = f_2(b) = e$, $f_2(c) = d$.
- Хотим: $f_1(L_1 \cap L_2) \cap f_2(L_3 \cap L_4) \in [[[\mathcal{L}]_{\mathrm{LFRel}}]_\cap]_{\mathrm{Hom}}$.
- $|f_1^{-1}(e)| = |f_2^{-1}(e)| = 2$, $|f_1^{-1}(d)| = |f_2^{-1}(d)| = 1$.
- Выделяем в Σ $2 \times 2 = 4$ символа: e_{ba}, e_{bb}, e_{ca}, e_{cb}. Для порядка добавляем еще символ d_{ac}.
- $L_1 \cap L_2 = L_1' \Rightarrow$
  $L_1(d_{ac}, \{e_{ba}, e_{bb}\}, \{e_{ca}, e_{cb}\}) \cap L_2(d_{ac}, \{e_{ba}, e_{bb}\}, \{e_{ca}, e_{cb}\}) =$
  $L_1'(d_{ac}, \{e_{ba}, e_{bb}\}, \{e_{ca}, e_{cb}\})$

Иллюстрация доказательства

- Пусть даны четыре языка $L_i = L_i(a, b, c)$, $i = 1, 2, 3, 4$ из класса $[\mathcal{L}]_{\mathrm{LFRel}}$, слова в которых состоят из символов a, b, c.
- Пусть $f_1, f_2 : \Sigma \to \Sigma$ задают символьные гомоморфизмы, причём
  1. $f_1(a) = d$, $f_1(b) = f_1(c) = e$;
  2. $f_2(a) = f_2(b) = e$, $f_2(c) = d$.
- Хотим: $f_1(L_1 \cap L_2) \cap f_2(L_3 \cap L_4) \in [[[\mathcal{L}]_{\mathrm{LFRel}}]_\cap]_{\mathrm{Hom}}$.
- $|f_1^{-1}(e)| = |f_2^{-1}(e)| = 2$, $|f_1^{-1}(d)| = |f_2^{-1}(d)| = 1$.
- Выделяем в Σ $2 \times 2 = 4$ символа: e_{ba}, e_{bb}, e_{ca}, e_{cb}. Для порядка добавляем еще символ d_{ac}.
- $L_1 \cap L_2 = L'_1 \Rightarrow$
  $L_1(d_{ac}, \{e_{ba}, e_{bb}\}, \{e_{ca}, e_{cb}\}) \cap L_2(d_{ac}, \{e_{ba}, e_{bb}\}, \{e_{ca}, e_{cb}\}) =$
  $L'_1(d_{ac}, \{e_{ba}, e_{bb}\}, \{e_{ca}, e_{cb}\})$
- Аналогично, $L_3 \cap L_4 = L'_2 \Rightarrow$
  $L_3(\{e_{ba}, e_{ca}\}, \{e_{bb}, e_{cb}\}, d_{ac}) \cap L_4(\{e_{ba}, e_{ca}\}, \{e_{bb}, e_{cb}\}, d_{ac}) =$
  $L'_2(\{e_{ba}, e_{ca}\}, \{e_{bb}, e_{cb}\}, d_{ac})$

Иллюстрация доказательства

- Определим гомоморфизм π так: $\pi(d_{ac}) = d$, $\pi(e_{xy}) = e$.

Иллюстрация доказательства

- Определим гомоморфизм π так: $\pi(d_{ac}) = d$, $\pi(e_{xy}) = e$.
- Итого:

$$\pi(L_1(d_{ac}, \{e_{ba}, e_{bb}\}, \{e_{ca}, e_{cb}\}) \cap L_2(d_{ac}, \{e_{ba}, e_{bb}\}, \{e_{ca}, e_{cb}\}) \cap L_3(\{e_{ba}, e_{ca}\}, \{e_{bb}, e_{cb}\}, d_{ac}) \cap L_4(\{e_{ba}, e_{ca}\}, \{e_{bb}, e_{cb}\}, d_{ac})) = f_1(L_1 \cap L_2) \cap f_2(L_3 \cap L_4)$$

Следствия и обобщения

- $\mathcal{CFL}$ — контекстно-свободные языки $\Rightarrow [[\mathcal{CFL}]_\cap]_{\mathrm{Hom}}$ замкнут относительно пересечений.

Следствия и обобщения

- $\mathcal{CFL}$ — контекстно-свободные языки $\Rightarrow [[\mathcal{CFL}]_\cap]_{\mathrm{Hom}}$ замкнут относительно пересечений.
- Доказательство основной теоремы может быть обобщено на произвольные локальные мощности алфавитов, символьных соответствий и мощности пересечений (нужно только следить, чтобы мощность множества вводимых символов не превосходила мощности алфавита; из-за этого придется требовать выполнения неравенств на кардиналы).

Следствия и обобщения

- $\mathcal{CFL}$ — контекстно-свободные языки $\Rightarrow [[\mathcal{CFL}]_\cap]_{\mathrm{Hom}}$ замкнут относительно пересечений.
- Доказательство основной теоремы может быть обобщено на произвольные локальные мощности алфавитов, символьных соответствий и мощности пересечений (нужно только следить, чтобы мощность множества вводимых символов не превосходила мощности алфавита; из-за этого придется требовать выполнения неравенств на кардиналы).
- Доказательство может быть обобщено на графовые языки.

В терминах категориальных грамматик

Результат может быть переформулирован в терминах категориальных грамматик: для этого в качестве $\mathcal{L}$ следует взять класс языков, задаваемых категориальными грамматиками над некоторым формализмом (например, над исчислением Ламбека).

Теорема

Существует полиномиальный алгоритм, которому на вход даются 2 набора категориальных грамматик Ламбека $Gr_1, \ldots, Gr_n$ и $Gr'_1, \ldots, Gr'_m$ и символьные гомоморфизмы $f, g : \Sigma \to \Sigma$ и который по ним строит такие грамматики $\widetilde{Gr}_1, \ldots, \widetilde{Gr}_k$ и символьный гомоморфизм π, что

$$f\left(L(Gr_1) \cap \cdots \cap L(Gr_n)\right) \cap g\left(L(Gr'_1) \cap \cdots \cap L(Gr'_m)\right) = \\ = \pi\left(L(\widetilde{Gr}_1) \cap \cdots \cap L(\widetilde{Gr}_k)\right)$$

О конъюнкции в категориальных грамматиках

Теорема

Языки, задаваемые грамматиками исчисления Ламбека с конъюнкцией на верхнем уровне (типы в нем устроены как $T_1 \wedge \cdots \wedge T_n$*, где* T_i *без конъюнкции), суть в точности символьные гомоморфизмы конечных пересечений контекстно-свободных языков без пустого слова.*

О конъюнкции в категориальных грамматиках

Теорема

Языки, задаваемые грамматиками исчисления Ламбека с конъюнкцией на верхнем уровне (типы в нем устроены как $T_1 \wedge \dots \wedge T_n$, где T_i без конъюнкции), суть в точности символьные гомоморфизмы конечных пересечений контекстно-свободных языков без пустого слова.

Пример

Из примера Пэуна (1980) следует, что в исчислении Ламбека с конъюнкцией на верхнем уровне можно задать язык $\{a^{2n^2} \mid n > 0\}$ и $\{a^{4k^2+2} \mid k > 0\}$. Тогда их пересечение также можно задать в данном варианте исчисления Ламбека. ***Получится язык*** $\{a^{\lceil \frac{1}{2}(3+2\sqrt{2})^{2n} \rceil + 1}\}$**.**

О локально конечных символьных соответствиях

- В работе показано, что условие замкнутости относительно локально конечных символьных соответствий на класс языков $\mathcal{L}$ равносильно тому, что $\mathcal{L}$ есть класс языков, задаваемых категориальными грамматиками над некоторым формализмом (в общем смысле; для этого было введено понятие формализма).

О локально конечных символьных соответствиях

- В работе показано, что условие замкнутости относительно локально конечных символьных соответствий на класс языков $\mathcal{L}$ равносильно тому, что $\mathcal{L}$ есть класс языков, задаваемых категориальными грамматиками над некоторым формализмом (в общем смысле; для этого было введено понятие формализма).
- Построен пример, показывающий, что $[[[\mathcal{L}]_{\mathrm{Hom}}]_{\cap}]_{\mathrm{Hom}}$ не обязательно замкнут относительно пересечений.

Спасибо за внимание!