Секция «Математическая логика, алгебра и теория чисел»

Последовательное замыкание классов языков относительно пересечения и гомоморфизмов

Научный руководитель – Пентус Мати Рейнович

Пшеницын Тихон Григорьевич
Студент (специалист)
Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Москва, Россия
E-mail: ptihon@yandex.ru

Одна из изучаемых тем в теории формальных языков — это замкнутость тех или иных классов языков относительно различных теоретико-множественных операций. В других наших исследованиях нередко возникал следующий вопрос: как по данному классу формальных языков понять, (а) замкнут ли он относительно пересечений, (б) что за класс получится, если замкнуть данный относительно пересечения? Мы изучали этот вопрос в контексте категориальных грамматик, языки которых по умолчанию замкнуты относительно символьных гомоморфизмов. Из-за этого возник ещё один вопрос: если последовательно замыкать класс языков относительно пересечения и гомоморфизмов, то после какого шага мы перестанем получать новые классы? На данный вопрос мы отвечаем в настоящей работе.

Сперва мы формулируем общее понятие категориальной грамматики. Для этого вводится серия таких определений:

Определение 1. Формализм — это структура $\langle Tp, g\rangle$, включающая в себя множество Tp и функцию $g : Tp^* \times Tp \to \{0, 1\}$.

Определение 2. Категориальная грамматика Gr над $\langle Tp, g\rangle$ — это структура вида $Gr = \langle \Sigma, S, \triangleright\rangle$, где Σ — не более чем счётный алфавит, $S \in Tp$ — выделенный тип, $\triangleright \subseteq \Sigma \times Tp$ — конечное бинарное отношение.

Определение 3. Язык, задаваемый (порождаемый) категориальной грамматикой $Gr = \langle \Sigma, S, \triangleright\rangle$ над формализмом $\langle Tp, g\rangle$ — это множество

$$L(Gr) = \{a_1 \dots a_n \in \Sigma^* : \exists T_1, \dots, T_n \in Tp : a_i \triangleright T_i, g(T_1, \dots, T_n; S) = 1\}.$$

Определение 4. Класс языков $\mathcal{L}$ над одним и тем же алфавитом Σ 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 $f(a_1 \dots a_n) := f(a_1) \dots f(a_n)$ (и $f(\varepsilon) = \varepsilon$), а на множества — по правилу $f(L) := \{f(w) | w \in L\}$.

Оказывается, что имеет место следующая

Теорема 1. *Пусть $\mathcal{L}$ — класс языков, порождаемых категориальными грамматиками над данным формализмом $\langle Tp, g\rangle$* [1]. *Тогда если замкнуть этот класс сначала относительно пересечений, а потом относительно символьных гомоморфизмов, то получится класс, замкнутый относительно пересечений.*

[1] В первоначальной формулировке мы накладываем на этот формализм дополнительное требование, которое впоследствии устраняем.

Доказательство теоремы конструктивно; оно даёт нам алгоритм, который по грамматикам Gr^i_j и символьным гомоморфизмам f^i строит такие грамматики Gr'_k и гомоморфизм π, что

$$\bigcap_{i=1}^{N} f^i \left(\bigcap_{j=1}^{M} L(Gr^i_j) \right) = \pi(L(Gr'_1) \cap \cdots \cap L(Gr'_k)).$$

Важная промежуточная конструкция при доказательстве теоремы — введение в формализм конъюнкции. Из доказательства теоремы выводится информация о том, какие в точности классы языков порождают такие формализмы с конъюнкцией (что даёт, в частности, точное описание языков, задаваемых групповыми грамматиками с конъюнкцией, исчислением Ламбека с конъюнкцией на верхнем уровне [1] и другими такого рода формализмами).

Помимо этого, созданное нами доказательство нигде по существу не использует то, что мы работаем со строковыми языками. Если вместо строковых языков рассмотреть гиперграфовые языки, то можно аналогично определить понятие формализма, категориальной грамматики и языка, задаваемого грамматикой. Тогда, в частности, результат Теоремы 1 будет верен и для класса гиперграфовых языков, задаваемых грамматиками замещения гиперрёбер (которые могут быть сведены к некоторому классу категориальных гиперграфовых грамматик; см. Pshenitsyn 2020).

Наконец, в работе мы изучаем, какие в принципе классы языков являются классами, порождаемыми грамматиками над некоторым формализмом. Мы установили, что необходимое и достаточное условие для этого — замкнутость относительно так называемых символьных соответствий. Мы также построили пример, когда незамкнутый относительно символьных соответствий класс утверждению Теоремы 1 не удовлетворяет.

Источники и литература

1) Kanazawa, M. 1992. The Lambek calculus enriched with additional connectives. Journal of Logic, Language and Information 1, 141–171.
2) Pshenitsyn, T. 2020. Hypergraph Basic Categorial Grammars. In: Gadducci F., Kehrer T. (eds) Graph Transformation. ICGT 2020. Lecture Notes in Computer Science, vol 12150. Springer, Cham. https://doi.org/10.1007/978-3-030-51372-6_9

[1] Исчисление Ламбека с конъюнкцией введено в Kanazawa 1992.