Мальцевские чтения
Новосибирск (2006)
Тезисы докладов

О распознавании языков полугруппами и автоматами

В. А. Молчанов

Работа посвящена обобщению классической теории формальных языков [1] на языки произвольных слов (содержащие как конечные, так и бесконечные в любую сторону слова).

Рассмотрим конечный алфавит A. Пусть $W_{fin}(A)$ – множество всех конечных слов, $W^{\rightarrow}(A)$ – множество всех бесконечных вправо слов, $W^{\leftarrow}(A)$ – множество всех бесконечных влево слов, $W^{\leftrightarrow}(A)$ – множество всех бесконечных в обе стороны слов и $W(A) = W_{fin}(A) \cup W^{\rightarrow}(A) \cup W^{\leftarrow}(A) \cup W^{\leftrightarrow}(A)$ – множество всех слов над алфавитом A. Подмножества $W(A)$ называются языками над алфавитом A.

В работе [2] с помощью методов нестандартного анализа [3] описаны все языки $L \subset W(A)$, которые распознаются автоматами Буши.

Ясно, что множество всех слов $W(A)$ можно рассматривать как четырехосновную алгебру $W(A) = (W_{fin}(A), W^{\leftarrow}(A), W^{\rightarrow}(A), W^{\leftrightarrow}(A))$ с канонически определенными в ней четырьмя операциями конкатенации подходящих слов и двумя унарными операциями над конечными словами $u \in W_{fin}$, которые определяются по формулам: $u^{+\omega} = uu:$, $u^{-\omega} =: uu$. При этом для любых $x, y, z \in W_{fin}(A)$, $u \in W^{\leftarrow}(A)$, $v \in W^{\rightarrow}(A)$ и натурального числа n выполняются свойства: $(xy)z = x(yz)$, $u(xy) = (ux)y$, $(xy)v = x(yv)$, $(ux)v = u(xv)$, $x(yx)^{+\omega} = (xy)^{+\omega}$, $(x^n)^{+\omega} = x^{+\omega}$, $(yx)^{-\omega}y = (xy)^{-\omega}$, $(x^n)^{-\omega} = x^{-\omega}$. Удовлетворяющие таким свойствам четырехосновные алгебры называются алгебрами Уилки.

В работе [4] с помощью методов нестандартного анализа построен функтор, который каждой конечной полугруппе S ставит в соответствие конечную алгебру Уилки $\overline{S} = (S, S^{\leftarrow}, S^{\rightarrow}, S^{\leftrightarrow})$, так что для любого отображения φ алфавита A в полугруппу S существует однозначно определенный гомоморфизм $\overline{\varphi}$ алгебры слов $W(A)$ в алгебру $\overline{S}$, ограничение которого на множестве A совпадает с отображением φ. Будем говорить, что язык $L \subset W(A)$ распознается полугруппой S, если $L = \overset{-1}{\overline{\varphi}}(P)$ для некоторого подмножества $P \subset \overline{S}$.

ТЕОРЕМА 1. Каждый распознаваемый автоматом Буши язык $L \subset W(A)$ распознаваем полугруппой.

ТЕОРЕМА 2. Язык $L \subset W(A)$ в том и только том случае распознаваем полугруппой, если он является конечным объединением языков вида $X^{-\omega}Y$, $XY^{+\omega}$, $X^{-\omega}YZ^{+\omega}$, где X, Y, Z – рациональные языки [1] и $X^{+\omega} = \{u_1u_2\ldots : u_1, u_2, \dots \in X\}$, $X^{-\omega} = \{\ldots u_{-2}u_{-1} : u_{-1}, u_{-2}, \dots \in X\}$.

17 октября 2006

Так как при распознавании языков произвольных слов автоматы Буши не являются алгебраическими эквивалентами конечных полугрупп, то рассматриваются также обобщенные автоматы Мюллера [1], которые представляются в виде алгебраических систем $\mathcal{A} = (Q, A, E, c, F, \mathcal{I}, \mathcal{F}, \mathcal{T})$, где:

(1) (Q, A, E) – автомат [1] с множеством состояний Q, входным алфавитом A и множеством переходов $E \subset Q \times A \times Q$,
(2) c – элемент множества Q, называемый центральным состоянием,
(3) F – подмножество множества Q, называемое множеством заключительных состояний,
(4) $\mathcal{I}$ и $\mathcal{F}$ – подмножества множества Q, называемые соответственно левой и правой таблицами состояний,
(5) $\mathcal{T}$ – множество упорядоченных пар подмножеств множества Q, называемое таблицей состояний.

Для такого автомата $\mathcal{A}$ обычным образом вводятся понятия пути (как конечной или бесконечной в любую сторону последовательности последовательных переходов автомата), метки пути, начала и конца пути. Если p – бесконечный путь в автомате $\mathcal{A}$, проходящий через центральное состояние c, то $Q_{-\infty}(p)$ (соответственно, $Q_{+\infty}(p)$) обозначает множество состояний, которые путь p посещает бесконечно много раз до (соответственно, после) посещения состояния c. Путь p в автомате $\mathcal{A}$ называется успешным, если он проходит через центральное состояние c и удовлетворяет одному из следующих условий:

(1) p есть конечный путь с началом c и концом в F,
(2) p есть такой бесконечный влево (соответственно, вправо) путь, что множество $Q_{-\infty}(p)$ (соответственно, $Q_{+\infty}(p)$) принадлежит таблице $\mathcal{I}$ (соответственно, таблице $\mathcal{F}$),
(3) p есть такой бесконечный в обе стороны путь, что упорядоченная пара множеств $(Q_{-\infty}(p), Q_{+\infty}(p))$ принадлежит таблице $\mathcal{T}$.

Множество меток всех успешных путей в автомате $\mathcal{A}$ обозначается символом $L(\mathcal{A})$. Будем говорить, что язык $L \subset W(A)$ распознается автоматом Мюллера, если $L = L(\mathcal{A})$ для некоторого обобщенного автомата Мюллера $\mathcal{A}$.

ТЕОРЕМА 3. Язык $L \subset W(A)$ в том и только том случае распознается полугруппой, если он распознается автоматом Мюллера.

Список литературы

[1] J. E. Pin, *Finite semigroups and recognizable languages: an introduction*, In book: Semigroups, Formal Languages and Groups, NATO ASI Series C: Mathematical and Physical Sciences, **466** (1993). P. 1–32.

[2] V. A. Molchanov, *Nonstandard approach to general rational languages*, Contributions to General Algebra**13**(2001). P. 233–244.

[3] С. Альбеверио , Й. Фенстад , Р. Хеэг-Крон , Т. Линдстрем, *Нестандартные методы в стохастическом анализе и математической физике*. М.: Мир. 1990.

[4] V. A. Molchanov, *О естественном продолжении теории рациональных языков на языки произвольных слов*// Математика, механика: Сб. науч. тр. Вып. 6. Саратов: Изд-во Сарат. ун-та. 2004. С. 90–93.

E-mail: v.molchanov@inbox.ru