Представление закодированных регулярных событий в абстрактных конечных автоматах *

Н. В. Крылов

Пусть дана свободная полугруппа с единицей F.

Определения

1. *Событием* $S(x_1, \ldots, x_n)$ в полугруппе $F(x_1, \ldots, x_n)$ называется любое множество слов, принадлежащих $F(x_1, \ldots, x_n)$.
2. *Итерацией* множества слов M называется операция образования минимальной полугруппы с единицей, содержащей M.

* С несколько иной точки зрения этот вопрос рассматривается в работе В. Г. Боднарчука (УМЖ, т. XIV, № 4, 1962).

3. Событие называется *регулярным*, если его множество слов может быть получено применением операций умножения, дизъюнкции (объединения) и итерации в конечном числе к алфавиту $x_1, \ldots, x_n$.

4. Представление регулярного события через $x_1, \ldots, x_n$ называется *регулярным выражением* или *формулой*.

5. Конечным абстрактным автоматом $A(a_0, \ldots, a_k; \bar{x}_1, \ldots, \bar{x}_n)$ называется конечное абстрактное множество элементов $a_0, \ldots, a_k$ (состояния автомата) с определенными в нем отображениями в себя $\bar{x}_1, \ldots, \bar{x}_n$ (входы автомата) и выделенным состоянием a_0 (начальное состояние).

Обычным образом определяется произведение преобразований.

6. Автомат $A(a_0, \ldots, a_k; \bar{x}_1, \ldots, \bar{x}_l)$ *представляет* события $S_i(x_1, \ldots, x_n)$, $i = 1, \ldots, j$, если:

1) определено взаимно однозначное соответствие между $\bar{x}_{\alpha_1}, \ldots, \bar{x}_{\alpha_n}$ ($\alpha_i \neq \alpha_j$ при $i \neq j$) и $x_1, \ldots, x_n$:

$$x_i \longleftrightarrow \bar{x}_{\alpha_i};$$

2) в A могут быть выделены такие подмножества S'_i $i = 1, \ldots, j$, что при замене букв любого слова из S_i соответствующими значками преобразований элемент a_0 полученным преобразованием переводится в $a' \in S'_i$, и наоборот, т. е. если некоторое преобразование $\bar{x}_{\alpha_i}, \ldots, \bar{x}_{\alpha_k}$ переводит a_0 в $a' \in S'_m$, то $x_i \ldots x_k \in S_m$.

В этом случае также говорят, что S'_i представляет событие S_i.

7. Гомоморфизм φ свободной полугруппы $F(x_1, \ldots, x_n)$ в свободную полугруппу $\Phi(a_1, \ldots, a_l)$ называется *кодированием*.

Скажем, что кодирование *унитарно*, если соотношение $\varphi(x_i) = \varphi(x_j) a_{\alpha_i} \ldots a_{\alpha_m}$ невозможно ни для какого (в том числе и пустого) слова.

Предложение. *Если кодирование $\varphi: F \to \Phi$ унитарно, то φF — свободная полугруппа относительно системы образующих $\varphi x_1, \ldots, \varphi x_n$.*

Для доказательства установим, что φ — изоморфизм, для чего достаточно показать, что любому слову $a \in \varphi F$ соответствует единственный прообраз в F при гомоморфизме φ.

В самом деле, пусть $\varphi(x_{l_1}, \ldots, x_{l_j}) = \varphi(x_{r_1}, \ldots, x_{r_m}) = a_{i_1}, \ldots, a_{i_k}$, тогда

$$\varphi(x_{l_1}), \ldots, \varphi(x_{l_j}) = \varphi(x_{r_1}), \ldots, \varphi(x_{r_m}). \quad (*)$$

Допустим, что слова $\varphi(x_{r_1})$ и $\varphi(x_{l_1})$ содержат L_1 и L'_1 букв соответственно (при этом пустое слово за букву не считается).

Могут представиться лишь три возможности:

$$1.\ L_1 < L'_1 \qquad 2.\ L'_1 < L_1 \qquad 3.\ L_1 = L'_1.$$

Равенство (*) выполнено в свободной полугруппе Φ. Следовательно, выбрасывая пустое слово, получаем побуквенное совпадение равных слов.

В первых двух случаях одно из $\varphi(x_{l_1})$ и $\varphi(x_{r_1})$ есть левый делитель другого, что противоречит унитарности кодирования. В третьем случае имеем $\varphi(x_{l_1}) = \varphi(x_{r_1})$ или, в силу определения унитарности, $x_{l_1} = x_{r_1}$.

Аналогично рассуждая, получим $x_{l_2} = x_{r_2}$ и т. д. Отсюда $x_{l_1}, \ldots, x_{l_j} = x_{r_1}, \ldots, x_{r_m}$, что и требовалось.

8. Важными и интересными являются следующие

Теоремы:

1. *Конечный автомат может представить только регулярные события* ([1], теорема 2).

2. *Любое конечное число регулярных событий может быть представлено в одном конечном автомате* ([2], теорема 1).

Число регулярных событий может быть велико, и сами записи этих событий могут быть сложными, в некотором смысле. Поэтому важно уметь строить сложные автоматы, представляющие регулярные события со сложными формулами, исходя из менее сложных автоматов.

Ниже приводится один из возможных способов такого построения.

9. Пусть автомат Φ представляет события $A_1, \ldots, A_k$ в полугруппе $F(x_1, \ldots, x_n)$; $\Phi = \Phi(S_0^\Phi, \ldots, S_m^\Phi;\ \bar{x}_1, \ldots, \bar{x}_n)$ в (дальнейшем $\bar{x}_i$ и x_i не будут различаться);

событие A_i представлено множеством $S'_i = \{S_{i_1}, \ldots, S_{i_{\alpha_i}}\}$, $i = 1, \ldots, k$, задано унитарное кодирование $\varphi\colon F \to G(a_1, \ldots, a_l)$ и $\varphi(x_j) = b_j = a_{j_1} \ldots a_{j_{\beta_j}}$, $j = 1, \ldots, n$;

далее, автомат $f(S_0^f, \ldots, S_p^f;\ a_1, \ldots, a_l)$ представляет события $b_1, \ldots, b_n$, причем b_i представлено состоянием $S_{k_i}^f$.

Очевидно, ни одна последовательность $a_{i_1}, \ldots, a_{i_k}$ не переводит S_0^f в S_0^f, значит, $k_i \neq 0$, для всех i.

Действительно, в противном случае можно было бы перевести автомат f в состояние $S_{k_i}^f$, действуя на S_0^f не только словом b_i, но и словом $a_{i_1} \ldots a_{i_k} b_i$, что противоречит определению представления b_i в f.

Построим теперь автомат $\psi = \Phi(f)$.

Для этого нужно определить состояния, входы, преобразования, которые производят входы над состояниями, и начальное состояние.

Состояниями автомата ψ будем считать классы пар (S_i^Φ, S_j^f), где $j \neq 0$. Пару $(S_i^\Phi S_j^f)$ отнесем к классу, обозначенному $(S_q^\Phi S_0^f)$, если S_j^f представляет событие b_γ и $S_i^\Phi x_\gamma = S_q^\Phi$. Если S_j^f не представляет ни одно из событий $b_1, \ldots, b_n$, то будем считать, что пара $(S_i^\Phi S_j^f)$ составляет самостоятельный класс.

Покажем, что классы пар не пересекаются. Пусть (S_i^Φ, S_j^f) принадлежит двум классам $(S_p^\Phi S_\alpha^f)$ и (S_q^Φ, S_β^f). Если состояние S_j^f представляет некоторое событие b_γ, то $S_\alpha^f = S_\beta^f = S_0^f$, а $S_i^\Phi x_\gamma = S_p^\Phi = S_q^\Phi$. В противном случае (S_i^Φ, S_j^f) принадлежит классу, состоящему из одной пары $(S_i^\Phi S_j^f)$, т. е. $(S_i^\Phi, S_j^f) = (S_p^\Phi S_\alpha^f) = (S_q^\Phi S_\beta^f)$.

Входами автомата ψ будем считать алфавит $a_1, \ldots, a_l$. Действие входа a_k на класс (S_i^Φ, S_j^f) определим равенством: кл $(S_i^\Phi S_j^f)\, a_k =$ кл $(S_i^\Phi, S_j^f a_k)$.

Наконец, за начальное состояние автомата ψ примем $(S_0^\Phi S_0^f) = \psi_0$.

10. Теорема. *Состояния автомата ψ, представляющего $\varphi A_1, \ldots, \varphi A_n$, $(S_i^\Phi, S_0^f), \ldots, (S_{i_{\alpha_i}}^\Phi, S_0^f)$ представляют событие φA_i.*

Для доказательства достаточно показать, что под действием $a \in \varphi A_i$ состояние ψ_0 переходит в одно из $(S_{i_1}^\Phi S_0^f), \ldots, (S_{i_{\alpha_i}}^\Phi S_0^f)$ и что если под действием слова a состояние ψ_0 переходит в одно из $(S_{i_1}^\Phi S_0^f), \ldots, (S_{i_{\alpha_i}}^\Phi S_0^f)$, то $a \in \varphi A_i$.

Докажем первое.

Пусть $a_{i_1} \ldots a_{i_\gamma} \in \varphi A_i$, в силу унитарности кодирования, а значит, изоморфизма между φF и F, это слово единственным образом можно представить в виде $\varphi(x_{j_1} \ldots x_{j_\nu}) = b_{j_1} \ldots b_{j_\nu}$, где $x_{j_1} \ldots x_{j_\nu} \in A_i$.

Докажем, что при применении к $(S_k^\Phi S_0^f)$ входа b_i автомат ψ перейдет в состояние $(S_j^\Phi S_0^f)$, где $S_j^\Phi = S_k^\Phi x_i$. Пусть $b_i = a_{i_1} \ldots a_{i_{\alpha_i}}$. Имеем $(S_k^\Phi S_0^f)a_{i_1} =$ $=$ кл $(S_k^\Phi,\ S_0^f a_{i_1})$. Далее, из-за унитарности кодирования $a_{i_1} \neq b_r$ $(r=1 \ldots n)$, в силу того что f представляет b_i, получим $S_0^f a_{i_1} \neq S_{k_i}^f$, $i = 1, \ldots, n$; отсюда кл $(S_k^\Phi S_0^f a_{i_1}) = (S_k^\Phi S_0^f a_{i_1})$. Аналогично $(S_k^\Phi S_0^f a_{i_1})\, a_{i_2} = (S_k^\Phi S_0^f a_{i_1} a_{i_2})$ и т. д. Наконец, $(S_k^\Phi S_0^f)\, a_{i_1} \ldots a_{i_{\alpha_i}-1} = (S_k^\Phi S_0^f a_{i_1} \ldots a_{i_{\alpha_i}-1})$; далее, $(S_k^\Phi S_0^f a_{i_1} \ldots a_{i_{\alpha_i}-1}) a_{i_{\alpha_i}} =$ $=$ кл $(S_k^\Phi S_0^f b_i)$. Пользуясь тем, что $S_0^f b_i$ представляет b_i, получаем

$$\text{кл } (S_k^\Phi S_0^f b_i) = (S_k^\Phi x_i S_0^f) = (S_j^\Phi S_0^f).$$

Отсюда следует, что

$$(S_0^\Phi S_0^f)\, b_{j_1} = (S_0^\Phi x_{j_1} S_0^f),$$

$$(S_0^\Phi S_0^f)\, b_{j_1} b_{j_2} = (S_0^\Phi x_{j_1} S_0^f)\, b_{j_2} = (S_0^\Phi x_{j_1} x_{j_2} S_0^f),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

и

$$(S_0^\Phi S_0^f)\, b_{j_1} \ldots b_{j_\nu} = (S_0^\Phi S_0^f)\, a_{i_1} \ldots a_{i_\gamma} = (S_0^\Phi x_{j_1} \ldots x_{j_\nu} S_0^f) = (S^\Phi_{i_\mu} S_0^f),$$

где μ — одно из $1, 2, \ldots, \alpha_i$, так как $x_{j_1} \ldots x_{j_\nu} \in A_i$.

Итак, при применении к состоянию ψ_0 элемента из φA_i автомат ψ переходит в одно из состояний $(S^\Phi_{i_1} S_0^f), \ldots, (S^\Phi_{i_{\alpha_i}} S_0^f)$, чем и доказана первая часть утверждения.

Докажем его вторую часть.

Пусть под действием слова $a_{i_1} \ldots a_{i_\gamma}$ автомат ψ перешел из состояния ψ_0 в одно из состояний $(S^\Phi_{i_1} S_0^f), \ldots, (S^\Phi_{i_{\alpha_i}} S_0^f)$, например, для определенности в $(S^\Phi_{i_1} S_0^f)$. Отметим в траектории этого перехода

$$(S_0^\Phi S_0^f)\,(S^\Phi_{k_1} S^f_{j_1}) \ldots (S^\Phi_{k_\delta} S^f_{j_\delta})\,(S^\Phi_{i_1} S_0^f)$$

те классы, в обозначении которых на втором месте стоит S_0^f.

Докажем, что отрезок траектории между двумя соседними отмеченными классами однозначно определяет последовательность a_i-х, вызвавшую этот переход из одного отмеченного класса в другой, и что эта последовательность есть одно из b_r $(r = 1, \ldots, n)$.

Пусть этот отрезок есть $(S^\Phi_{r_1} S_0^f) \ldots (S_i^\Phi S_j^f)\,(S^\Phi_{r_2} S_0^f)$ и пусть он вызван воздействием последовательности $a_{k_1} \ldots a_{k_l}$ на состояние $(S^\Phi_{r_1} S_0^f)$.

Ввиду того что состояние S_0^f появляется в нем только на концах, имеем

$$(S^\Phi_{r_1} S_0^f)\, a_{k_1} \ldots a_{k_{l-1}} = (S^\Phi_{r_1} S_0^f a_{k_1} \ldots a_{k_{l-1}})$$

и

$$(S^\Phi_{r_2} S_0^f) = (S^\Phi_{r_1} S_0^f a_{k_1} \ldots a_{k_{l-1}})\, a_{k_l} = \text{кл } (S^\Phi_{r_1} S_0^f a_{k_1} \ldots a_{k_l});$$

как уже указывалось, $S_0^f a_{k_1} \ldots a_{k_l} \neq S_0^f$. Поэтому из последнего равенства следует, что $S_0^f a_{k_1} \ldots a_{k_l}$ представляет одно из b_r, а так как f представляет b_r, то $a_{k_1} \ldots a_{k_l} = b_r$. Кроме того,

$$S^\Phi_{r_1} x_r = S^\Phi_{r_2}.$$

Теперь, так как траектория целиком распадается на куски рассмотренного типа, то $a_{i_1} \ldots a_{i_\gamma} = b_{r_1} \ldots b_{r_\sigma}$ и $S_0^\Phi x_{r_1} \ldots x_{r_\sigma} = S^\Phi_{i_1}$. Поскольку автомат

Φ представляет событие A_i и $S^{\Phi}_{i_1} \in S_i'$, то $x_{r_1} \ldots x_{r_\sigma} \in A_i$. Отсюда же следует, что $b_{r_1} \ldots a_{r_\sigma} \in \varphi A_i$, что доказывает вторую часть утверждения.

Теорема доказана.

11. Определение. Назовем автомат ψ суперпозицией Φ и f с выделенными элементами $b_i = a_{i_1} \ldots a_{i_{\alpha_i}}$, $i = 1 \ldots n$.

В точности так же, как дается определение суперпозиции автоматов с выделенными элементами, дается определение суперпозиции Φ и f с выделенными регулярными событиями $S_1, \ldots, S_k$, для которых S_i' (состояния автомата f, представляющие S_i) не пересекаются. Именно, пусть автомат $f(S_0^f \ldots S_m^f;\ a_1, \ldots, a_l)$ представляет события $S_1, \ldots, S_k$, пусть S_i' не пересекаются и дан некоторый автомат $\Phi(S_0^{\Phi}, \ldots, S_n^{\Phi};\ x_1, \ldots, x_k)$.

Состояниями автомата ψ суперпозиции автоматов f и Φ будем считать классы пар $(S_i^{\Phi} S_j^f)$, где $j \neq 0$. Пару $(S_i^{\Phi} S_j^f)$ отнесем к классу $(S_p^{\Phi} S_0^f)$, если $S_j^f \in S_\gamma'$ при некотором γ и $S_i^{\Phi} x_\gamma = S_p^{\Phi}$. Если $S_j^f \bar{\in} S_\gamma'$ при всех γ, то будем считать, что пара $(S_i^{\Phi} S_j^f)$ составляет самостоятельный класс. Очевидно, классы пар не пересекаются.

Входами автомата ψ будем считать $a_1, \ldots, a_l$; действия входа a_k на состояние кл $(S_i^{\Phi} S_j^f)$ определим равенством кл $(S_i^{\Phi} S_j^f)\, a_k =$ кл $(S_i^{\Phi} S_j^f a_k)$.

За начальное состояние автомата ψ примем $(S_0^{\Phi} S_0^f)$.

Из доказанного в п. 10 вытекает способ построения автомата, представляющего события со сложными регулярными записями-формулами, через автоматы, представляющие события с менее сложными регулярными записями. Именно, пусть даны некоторые регулярные события. Обозначим в них некоторые подформулы буквами $X_1, \ldots, X_n$ и вставим эти обозначения в основные формулы, необходимо только, чтобы это обозначение было унитарным, т. е. чтобы ни одно слово события X_j не являлось левым делителем ни одного слова события X_i $(i \neq j)$ и $\in X_i$.

Теперь в любом автомате f, представляющем $X_1, \ldots, X_n$, множества состояний X_i' не пересекаются, поэтому из-за унитарности обозначений можно считать, что X_j' состоит из одного элемента-состояния.

Представим теперь $X_1 = (\ldots), \ldots, X_n = (\ldots)$ в таком автомате f и преобразованные события — в автомате Φ, затем сделаем суперпозицию ψ автоматов Φ и f с выделенными регулярными событиями $X_1, \ldots, X_n$.

Ясно, что автомат ψ представляет исходные события.

ЛИТЕРАТУРА

1. Ю. Т. Медведев, Сб. «Автоматы», ИЛ, М., 1956, 385—401.
2. В. М. Глушков, УМЖ, т. XII, № 2, 1960, 147.

Поступила 6.VII 1961 г.
Москва