О звездной высоте регулярного языка и циклической сложности минимального автомата

В. М. Дементьев

Понятие звездной высоты было впервые введено Эгганом в 1963 г. наряду с понятием циклической сложности конечного автомата. Циклическая сложность автомата связана со звездной высотой принимаемого им языка: звездная высота регулярного языка равна минимальной циклической сложности среди автоматов, принимающих данный язык. Однако проблема нахождения минимального (по отношению к циклической сложности) автомата до сих пор остается открытой. В данной работе рассматривается вопрос о соотношении циклической сложности минимального автомата регулярного языка и его звездной высоты.

Ключевые слова: конечные автоматы, регулярные языки, звездная высота, циклическая сложность, минимальный автомат.

1. Введение

Понятие звездной высоты регулярного языка было впервые введено Эгганом [2] в 1963 году как мера сложности регулярного множества. В своей работе Эгган поставил два вопроса: существует ли регулярные языки сколь угодно большой звездной высоты и существует ли алгоритм вычисления звездной высоты для произвольного регулярного множества.

На первый вопрос самим Эгганом был получен положительный ответ в случае алфавита нефиксированной длины. В случае фиксированного алфавита (в частности $\{0,1\}$) Дижоном и Шутценберже [3] в 1966 году был получен также положительный ответ.

Второй вопрос оставался открытым в течение 25 лет, до тех пор, пока Хашигучи [4] в 1988 году не предъявил алгоритм для вычисления звездной высоты регулярного выражения. Однако данный алгоритм не применим на практике в виду высокой вычислительной сложности (вычисление звездной высоты для языка, принимаемого автоматом с четырьмя состояниями, требует перебора $10^{\overbrace{\cdot^{\cdot^{\cdot}}}^{7}\,10}$ регулярных языков [5]).

Более эффективный алгоритм (дважды экспоненциальной сложности) был предложен Кирстеном [6] в 2005 году.

Кроме того Эгганом было введено понятие циклической сложности автомата (ориентированного графа). Циклическая сложность автомата связана со звездной высотой принимаемого им языка: звездная высота регулярного языка равна минимальной циклической сложности среди автоматов, принимающих данный язык.

В данной работе рассматривается вопрос о соотношении циклической сложности минимального автомата регулярного языка и его звездной высоты.

Автор выражает благодарность своему научному руководителю, кандидату физико-математических наук, Строгалову А. С.

2. Определения и используемые теоремы

Определение 1. *Конечным детерминированным автоматом* называется набор (A, Q, F, φ, q_0), где Q — конечное множество состояний, $F \subseteq Q$ — множество финальных состояний, q_0 — начальное состояние, $\varphi : A^* \times Q \to Q$, $\varphi(q, \lambda) = q$, $\varphi(q, \alpha a) = \varphi(\varphi(q, \alpha), a)$, где λ — пустой символ.

Каждому автомату $\mathcal{A}$ соответствует *регулярный язык* $\{w \in A^* \mid \varphi(q_0, w) \in F\}$. Будем обозначать этот язык $L(\mathcal{A})$ и говорить, что автомат $\mathcal{A}$ *принимает* язык L. Множество всех автоматов, принимающих L, обозначим $M(L)$.

Определение 2. *Минимальным автоматом* для языка L называется автомат, принимающий L и имеющий минимальное число состояний среди всех автоматов, принимающих L. Минимальный автомат определен однозначно с точностью до изоморфизма [1]. Будем обозначать минимальный автомат $\mathcal{A}_{min}(L)$.

Определение 3. *Регулярным выражением* над алфавитом A называется слово над алфавитом $A \cup \{\varnothing, \cdot, +, *, (,)\}$, определяемое по следующим правилам:

1) символы алфавита A и $\varnothing$ — регулярные выражения;
2) если α и β — регулярные выражения, то $(\alpha + \beta)$, $(\alpha \cdot \beta)$, $(\alpha)^*$ — регулярные выражения.

Обозначим регулярный язык, соответствующий регулярному выражению $\mathcal{R}$ как $L(\mathcal{R})$. Множество регулярных выражений, определяющих язык L, обозначим $R(L)$.

Определение 4. *Звездная высота* регулярного выражения $\mathrm{sh}(\mathcal{R})$ определяется следующим образом:

1) $\mathrm{sh}(a) = 0 \; \forall a \in A \cup \lambda$;
2) $\mathrm{sh}(\alpha + \beta) = \mathrm{sh}(\alpha \cdot \beta) = \max(\mathrm{sh}(\alpha), \mathrm{sh}(\beta))$;
3) $\mathrm{sh}(\alpha^*) = \mathrm{sh}(\alpha) + 1$.

Определение 5. *Звездная высота* регулярного языка L равна $\min_{r \in R(L)} \mathrm{sh}(r)$.

Определение 6. *Циклическая сложность* ориентированного графа $\mathcal{G}$ определяется следующим образом:

1) если граф $\mathcal{G}$ не содержит сильносвязных компонент или является изолированной вершиной, то $\mathrm{lc}(\mathcal{G}) = 0$;
2) $\mathrm{lc}(\mathcal{G}) = \max\{\mathrm{lc}(\mathcal{P}) \mid \mathcal{P} - \text{сильносвязная компонента графа } \mathcal{G}\}$, если $\mathcal{G}$ не сильносвязный;
3) $\mathrm{lc}(\mathcal{G}) = 1 + \min\{\mathrm{lc}(\mathcal{G} \setminus \{s\}) \mid s \in V(\mathcal{G})\}$, если $\mathcal{G}$ сильносвязный.

Теорема 1. *[2] Звездная высота регулярного языка равна минимуму циклической сложности среди всех автоматов, принимающих L:*

$$\mathrm{sh}(L) = \min\{\mathrm{lc}(\mathcal{A}) \mid \mathcal{A} \in M(L)\}.$$

3. Основное утверждение

Заметим, что теорема 1 позволяет свести задачу вычисления звездной высоты к задаче нахождения автомата, минимального относительно циклической сложности, принимающего данный язык.

Легко привести пример регулярного языка, минимальный автомат (по числу состояний) которого не является минимальным автоматом относительно циклической сложности.

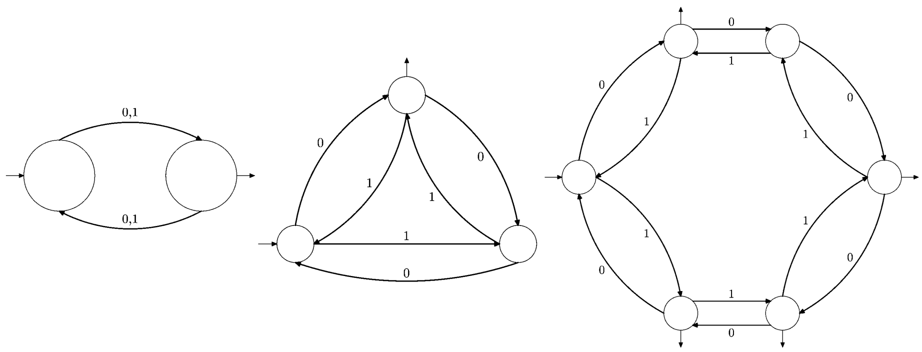

Рис. 1. Автоматы F_2, F_3 и H_6 (слева направо).

Пусть w^i равно числу символов i в слове $w \in A^*$. Рассмотрим языки $F_2 = \{w \in \{0{,}1\}^* \mid |w^0 - w^1| \equiv 1 \bmod 2$, $F_3 = \{w \in \{0{,}1\}^* \mid |w^0 - w^1| \equiv 1 \bmod 3$ и их объединение $H_6 = F_2 \cup F_3$. Соответствующие минимальные автоматы изображены на рисунке 1.

Циклические сложности автоматов F_2, F_3 и H_6 соответственно равны 1, 2 и 3. Звездная высота H_6 равна (по определению) максимуму из звездных высот F_2 и F_3, откуда следует, что она не превосходит 2. Таким образом минимальный автомат для H_6 не является минимальным относительно циклической сложности.

Возникает вопрос: существует ли связь между циклической сложностью минимального автомата и звездной высотой принимаемого им языка? Ответ на этот вопрос дает следующая

Теорема 2. *Разница между циклической сложностью регулярного языка и его звездной высотой, как функция от языка, неограничена:*

$$\forall\, n \in \mathbb{N} \;\; \exists L : \; \mathrm{lc}(\mathcal{A}_{min}(L)) - \mathrm{sh}(L) \geqslant n.$$

Доказательство. Далее считаем, что $A = \{0{,}1\}$.

1) Рассмотрим семейство языков $\{C_n\}_{n=2,3,\ldots}$ таких, что $\mathcal{A}_{min}(C_n)$ имеет $Q = \{q_i\}_{i=1}^{n+1}$, а функция переходов φ определна следующим образом:

$$\varphi(a, q_i) = \begin{cases} q_{i+1}, \text{ если } a = 0, \ i < n \\ q_{i-1}, \text{ если } a = 1, \ i > 0 \\ q_n, \text{ если } a = 1, \ i = 1 \text{ или } a = 0, \ i = n-1 \text{ или } i = n \end{cases}$$

Автомат $\mathcal{A}_{min}(C_n)$ изображен на рисунке 2. Он является минимальным, так как любые два состояния q_i, q_j, $i < j$ различимы словом $\underbrace{0 \ldots 0}_{(n-1)-j}$.

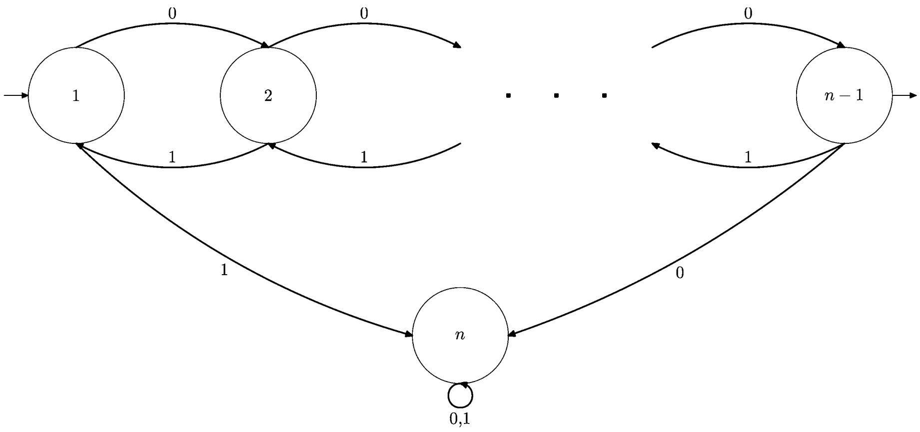

Рис. 2. Автомат $\mathcal{A}_{min}(C_n)$.

Лемма 1. *Циклическая сложность* $\mathrm{lc}(\mathcal{A}_{min}(C_n))$ *равна* $\lfloor \log_2 n \rfloor$.

Будем рассматривать диаграмму автомата $\mathcal{A}_{min}(C_n)$ без учета поглощающего состояния (его наличие не влияет на вычисление циклической сложности).
Заметим, что диаграмма автомата сильносвязна, следовательно его циклическая сложность равна

$$\mathrm{lc}(\mathcal{A}_{min}(C_n)) = 1 + \min \{\mathrm{lc}(\mathcal{A}_{min}(C_n)/s) \mid s \in V(\mathcal{A}_{min}(C_n))\}.$$

При отсечении k-й вершины диаграмма разбивается на две компоненты C_{k-1} и C_{n-k}.

Возможно два случая: минимум достигается, когда s — начальная или конечная вершина, и когда s — любая другая вершина. В первом случае имеем

$$\mathrm{lc}(\mathcal{A}_{min}(C_n)) = 1 + \mathrm{lc}(\mathcal{A}_{min}(C_{n-1})) > \mathrm{lc}(\mathcal{A}_{min}(C_{n-1})).$$

Рассмотрим второй случай. Тогда минимум достигается на компоненте C_{k-1} или C_{n-k} для некоторого $1 < k < n$. Будем считать, что это C_{k-1}. Но тогда данные компоненты входят и в вычисленеие $\mathrm{lc}(\mathcal{A}_{min}(C_{n-1}))$ для того же k, а значит $\mathrm{lc}(\mathcal{A}_{min}(C_n)) \geqslant \mathrm{lc}(\mathcal{A}_{min}(C_{n-1}))$.
Циклическая сложность $\mathcal{A}_{min}(C_n)$ вычисляется рекурсивно:

$$\mathrm{lc}(\mathcal{A}_{min}(C_1)) = 0, \;\; \mathrm{lc}(\mathcal{A}_{min}(C_2)) = \mathrm{lc}(\mathcal{A}_{min}(C_3)) = 1,$$
$$\mathrm{lc}(\mathcal{A}_{min}(C_n)) = 1 + \mathrm{lc}(\mathcal{A}_{min}(C_{\lfloor n/2 \rfloor})), \tag{1}$$

откуда получаем

$$\mathrm{lc}(\mathcal{A}_{min}(C_n)) = \lfloor \log_2 n \rfloor. \tag{2}$$

Лемма доказана.

2) Рассмотрим семейство языков $\{D_n\}_{n=2,3,\dots}$, где $D_n = \big\{w \in A^* \mid |w^0 - w^1| \equiv 1 \bmod n\big\}$.
Автомат $\mathcal{A}_{min}(D_n)$ изображен на рисунке 3. Он является минимальным, так как любые два состояния q_i, q_j, $i < j$ различимы словом $\underbrace{0\dots0}_{(n-1)-j}$.
Диаграмма данного автомата сильносвязна, следовательно для вычисления циклической сложности необходимо рассмотреть минимум по всевозможным диаграммам, которые получаются при отсечении одной вершины. В данном случае, независимо от вершины, будет получаться граф C_{n-1}. Следовательно

$$\mathrm{lc}(\mathcal{A}_{min}(D_n)) = 1 + \mathrm{lc}(\mathcal{A}_{min}(C_{n-1})) = 1 + \lfloor \log_2(n-1) \rfloor$$
$$\Rightarrow \mathrm{sh}(\mathcal{A}_{min}(D_n)) \leqslant 1 + \lfloor \log_2(n-1) \rfloor.$$

3) Для доказательства теоремы рассмотрим объединения языков D_n и D_m.
Рассмотрим автомат $\mathcal{A}_{min}(D_{nm})$. Добавим состояния $\{q_i \in Q \mid (i \equiv 1 \bmod n)$ или $(i \equiv 1 \bmod m)\}$ в множество финальных состояний. Тогда полученный автомат будет принимать язык $D_n \cup D_m$. Обозначим его $\mathcal{A}^+(D_{nm})$.

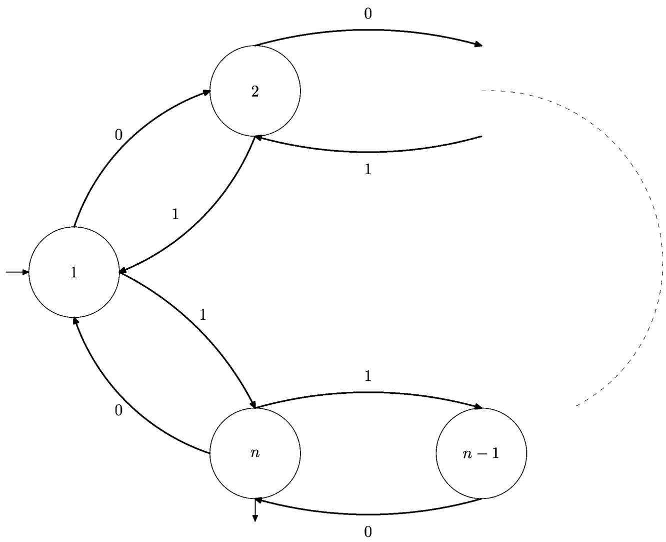

Рис. 3. Автомат $\mathcal{A}_{min}(D_n)$.

Лемма 2. *Автомат $\mathcal{A}^+(D_{pq})$ минимален для $D_p \cup D_q$, если p и q взаимно просты.*

Нам необходимо доказать, что в случае, когда p и q взаимно просты, все состояния автомата различимы.
Будем доказывать от противного. Пусть состояния q_i и q_j неразличимы.

а) $q_i,\ q_j \in F,\ i = kp+1,\ j = mq+1$. Тогда q_i и q_j различимы словом $\underbrace{0\ldots0}_{p}$ так как из q_i мы снова переходим в финальное состояние, а из состояния q_j в обычное состояние.

б) $q_i,\ q_j \in F,\ i = kp+1,\ j = mp+1$. Пусть $d = \min\{s \mid i+s \equiv 1 \bmod q\}$. Тогда $\forall n$ имеем

$$\begin{gathered} kp+1+d+nq \equiv 1 \bmod q, \\ mp+1+d+nq \equiv 1 \bmod q \\ \Rightarrow (k-m)p \equiv 1 \bmod q \Rightarrow k = m. \end{gathered}$$

Приходим к противоречию.

в) $q_i,\ q_j, i = kp+1,\ j = mq+1$. Пусть $d = \min\{s \mid q_{i+s} \in F\}$. Тогда либо $q_{j+s} \notin F$, либо попадаем в случай (а).

Лемма доказана.
Теперь перейдем непосредственно к доказательству.
Рассмотрим язык $D_p \cup D_q,\ p > q$. Имеем

$$\begin{gathered}\mathrm{lc}(\mathcal{A}_{min}(D_p \cup D_q)) = 1 + \lfloor \log_2(pq-1) \rfloor,\\ \mathrm{sh}(D_p \cup D_q) \leqslant 1 + \lfloor \log_2(p-1) \rfloor\\ \Rightarrow \mathrm{lc}(\mathcal{A}_{min}(D_p \cup D_q)) - \mathrm{sh}(D_p \cup D_q) \geqslant \log_2 \frac{pq-1}{p-1} =\\ = \log_2 \left(q + \frac{q}{p} + \frac{q}{p^2} + \ldots \right) > \log_2 q.\end{gathered}$$

Для любого n рассмотрим

$$\begin{gathered}p\ \text{ -- простое},\ p > 2^n,\\ q = 2^n\\ \Rightarrow \mathrm{lc}(\mathcal{A}_{min}(D_p \cup D_{2^n})) - \mathrm{sh}(D_p \cup D_{2^n}) \geqslant \log_2 2^n = n.\end{gathered}$$

Теорема доказана.

Список литературы

[1] Кудрявцев В. Б., Алешин С. В., Подколзин А. С. Введение в теорию автоматов. — М.: Наука, 1985.

[2] Eggan L. C. Transition graphs and the star-height of regular events // Michigan Math. J. — 1963. — 12. — Vol. 10, no. 4. — P. 385–397.

[3] Dejean F., Schützenberger M. P. On a Question of Eggan // Information and Control. — 1966. — February. — Vol. 9, no. 1. — P. 23–25.

[4] Hashiguchi K. Algorithms for Determining Relative Star Height and Star Height // Information and Computation. — 1988. — Vol. 78, no. 2. — P. 124–169.

[5] Lombardy S., Sakarovitch J. Star Height of Reversible Languages and Universal Automata // Proceedings of the 5th Latin American Symposium on Theoretical Informatics. — LATIN '02. — Springer-Verlag, 2002. — P. 76–90.

[6] Kirsten D. Distance desert automata and the star height one problem // FOSSACS'04 proceedings. — Springer-Verlag, 2004. — P. 2005.