7 Представление событий автоматами

7.1 Основные понятия

Подмножества M множества $A^* \setminus \{\Lambda\}$ называются *событиями* в алфавите A. Рассмотрим некоторый инициальный конечный автомат $V = (A, Q, B, \varphi, \psi, q)$ и некоторое подмножество $B' \subseteq B$. Тогда множество

$$M \subseteq A^*, M = \{\alpha | \psi(q, \alpha) \in B'\}$$

называется множеством слов (событием в алфавите A), *представимым в конечном автомате V с помощью подмножества B'* выходных символов. (Иначе: автомат V *представляет* событие M посредством B'). Событие M называется *представимым,* если существует конечный автомат V_q, представляющий это событие посредством некоторого подмножества B' выходных символов. Кроме обычных теоретико-множественных операций объединения, пересечения и дополнения будем использовать следующие операции:

1) *Произведение* событий M_1 и M_2

$$M_1 M_2 = \{\alpha_1 \alpha_2\} | \alpha_1 \in M_1, \alpha_2 \in M_2\};$$

2)*Итерация* события M :

$$M^* = \{\alpha_1 \alpha_2 \ldots \alpha_k\} | \alpha_i \in M, k \geq 1.$$

Эти операции обладают свойствами:

$$\emptyset M = M \emptyset = \emptyset,$$

$$\emptyset^* = \emptyset,$$

где $\emptyset = \{\}$ – пустое множество.

$$M^* = M M^* \cup M, M M^* = M^* M.$$

$$(A \cup B)C = AC \cup BC.$$

$$C(A \cup B) = CA \cup CB.$$

Событие $M, M \subseteq A^*$ называется *регулярным,* если его можно получить из событий вида $\emptyset, \{a\}, a \in A$ применением конечного числа операций объединения, произведения и итерации.

Например, регулярно любое конечное множество слов длины 1.

Регулярные события удобно представлять так называемыми *регулярными выражениями,* отражающими процесс построения регулярного события начиная с пустого и одноэлементных событий посредством операций произведения, объединения и итерации. Регулярные выражения представляют собой слова (формулы) в алфавите $A \cup \{\emptyset, \vee, \cdot, ^*, (,)\}$ и описываются следующим образом:

1) символы алфавита A и символ $\emptyset$ являются регулярными выражениями,

2) если α, β — регулярные выражения, то $(\alpha \vee \beta)$, $(\alpha \cdot \beta), \alpha^*$ также являются регулярными выражениями.

3) каждое регулярное выражение получается в соответствии с п.п. 1),2) данного определения за конечное числа шагов.

Обозначим R событие, соответствующее регулярному выражению $\mathcal{R}$, по определению,

$$R = \begin{cases} \emptyset \text{ если } \mathcal{R} = \emptyset, \\ \{a\} \text{ если } \mathcal{R} = a, \\ R_1 \cup R_2 \text{ если } \mathcal{R} = \mathcal{R}_1 \vee \mathcal{R}_2, \\ R_1 R_2 \text{ если } \mathcal{R} = \mathcal{R}_1 \cdot \mathcal{R}_2, \\ R_1^* \text{если } \mathcal{R} = \mathcal{R}_1^* \end{cases} .$$

Регулярные выражения называются эквивалентными (что обозначается $\mathcal{R}_1 \equiv \mathcal{R}_2$), если им соответствуют равные события.

Из свойства операций над событиями вытекают свойства регулярных выражений.

$$\emptyset \mathcal{R} \equiv \mathcal{R} \emptyset \equiv \emptyset,$$

$$\emptyset^* \equiv \emptyset,$$

$$\mathcal{R}^* = \mathcal{R}\mathcal{R}^* \vee \mathcal{R}, \mathcal{R}\mathcal{R}^* \equiv \mathcal{R}^*\mathcal{R}.$$

$$(\mathcal{R}_1 \vee \mathcal{R}_2)\mathcal{R}_3 \equiv \mathcal{R}_1\mathcal{R}_3 \vee \mathcal{R}_2\mathcal{R}_3.$$

$$\mathcal{R}_3(\mathcal{R}_1 \vee \mathcal{R}_2) \equiv \mathcal{R}_3\mathcal{R}_2 \vee \mathcal{R}_3\mathcal{R}_2.$$

Лемма 4.1. *Соотношение $X = XC \cup D$ выполняется для событий C, D и X тогда и только тогда, когда $X = DC^* \cup D$.*

Доказательство. Если $X = DC^* \cup D$, то

$$XC \cup D = (DC^* \cup D)C \cup D =$$

$$= (DC^*C \cup DC) \cup D = D(C^*C \cup C) \cup D = DC^* \cup D = X.$$

Пусть теперь $X = XC \cup D$. Тогда $X \subseteq DC^* \cup D$, поскольку в противном случае в $X \setminus (DC^* \cup D)$ найдется слово $\alpha \in DC^*$, противоречие.

(Этим словом будет кратчайшее слово α из $X \setminus (DC^* \cup D)$, при этом $\alpha \in XC \cup D, \alpha \notin D$ и поэтому $\alpha = \alpha_1\alpha_2 \in XC, \alpha_1 \in X, \alpha_2 \in C$, причем $\alpha_1 \notin (X \setminus DC^* \cup D)$, так как оно короче слова α — кратчайшего из этого множества и следовательно $\alpha_1 \in DC^* \cup D$ и $\alpha \in DC^*$:

$$\alpha_1 \in DC^* \cup D \rightarrow \alpha_1\alpha_2 \in DC^*C \cup DC = D(C^*C \cup C) = DC^*.)$$

Далее, если $X = XC \cup D$, то $DC^* \cup D \subseteq X$. В противном случае в $(DC^* \cup D) \setminus X$ найдется слово $\alpha \in X$, противоречие.

(Этим словом будет кратчайшее слово $\alpha \in (DC^* \cup D) \setminus X$, при этом $\alpha \notin X$ и, следовательно, $\alpha \notin D$, поэтому $\alpha = \alpha_1\alpha_2 \in DC^*$ ($\alpha \notin D$!), откуда ($\alpha_1 \in D$ или $\alpha_1 \in DC^*$) $\alpha_1 \in (DC^* \cup D)$, но $\alpha_1 \notin (DC^* \cup D) \setminus X$, так как оно короче самого короткого слова α из этого множества. Таким образом, $\alpha_1 \in X$ и, следовательно, $\alpha_2 \in C$. Значит $\alpha = \alpha_1\alpha_2 \in XC \subseteq X$. Отсюда $\alpha \in X$.)

Лемма доказана.

Лемма 4.2. *Пусть $R_{ij}, i = 0, 1, \ldots, n,\ \ j = 1, \ldots, n$ — регулярные события, $X_1, \ldots X_n$ — события, удовлетворяющие системе уравнений*

$$\begin{cases} X_1 = X_1R_{11} \cup \ldots \cup X_nR_{n1} \cup R_{01}, \\ \ldots\ldots\ldots\ldots\ldots \\ X_n = X_1R_{1n} \cup \ldots \cup X_nR_{nn} \cup R_{0n}. \end{cases}$$

Тогда события $X_1, \ldots X_n$ регулярны.

Доказательство. (Индукция по n.) В случае $n = 1$ $X_1 = X_1R_{11} \cup R_{01}$ и по лемме 4.1 $X_1 = R_{01}R_{11}^* \cup R_{01}$, то есть X_1 регулярно. Пусть утверждение доказано для некоторого $n = k - 1$, рассмотрим случай $n = k$. Из последнего уравнения системы по лемме 1 (полагая $C = R_{nn}$) получаем

$$X_n = (X_1R_{1n} \cup \ldots \cup X_{n-1}R_{n-1,n} \cup R_{0n})R_{nn}^* \cup$$

$$\cup X_1R_{1n} \cup \ldots \cup X_{n-1}R_{n-1,n} \cup R_{0n} =$$

$$= X_1(R_{1n}R_{nn}^* \cup R_{1n}) \cup \ldots \cup X_{n-1}(R_{n-1,n}R_{nn}^* \cup R_{n-1,n}) \cup$$

$$\cup R_{0n}R_{nn}^* \cup R_{0n}.$$

Подставив полученное выражение в первые n уравнений после приведения подобных членов получим систему того же вида для $X_1, \ldots X_{n-1}$. По предположению индукции события $X_1, \ldots, X_{n-1}$ оказываются регулярными, а вместе с ними регулярно и событие X_n, представленное выше через них с помощью операций объединения, произведения и итерации. Лемма доказана.

7.2 Теорема Клини

7.2.1 Регулярность событий, представимых конечными автоматами

Лемма 4.3. *Каждое событие, представимое в конечном автомате регулярно.*

Доказательство. Пусть событие M представимо в конечном автомате $V_{q_1} = (A, Q, B, \varphi, \psi, q_1), Q = \{q_1, \ldots, q_n\}$ с помощью множества $B', B' \subseteq B$:

$$M = \{\alpha | \alpha \in A^*, \psi(q_1, \alpha) \in B'\}.$$

Обозначим

$$M_i = \{\alpha | \alpha \in A^*, \alpha \neq \Lambda, \varphi(q_1, \alpha) = q_i\},$$
$$M'_i = \{a | a \in A, \psi(q_i, a) \in B'\}.$$

Очевидно, что события M'_i регулярны, а событие M представимо выражением

$$M = M_1 M'_1 \cup \ldots \cup M_n M'_n \cup M'_1$$

и для доказательства его регулярности достаточно убедиться в регулярности событий $M_1, \ldots M_n$.

Последние соответствуют системе уравнений

$$\begin{cases} M_1 = M_1 R_{11} \cup \ldots \cup M_n R_{n1} \cup R_{11}, \\ \quad\quad \ldots\ldots\ldots\ldots \\ M_n = M_1 R_{1n} \cup \ldots \cup M_n R_{nn} \cup R_{1n}, \end{cases},$$

где

$$R_{ij} = \{a | a \in A, \varphi(q_i, a) = q_j\}.$$

Учитывая регулярность событий R_{ij} и Лемму 4.2, получаем, что события $M_i, i = 1, \ldots n$, а вместе с ними и событие M регулярны. Лемма доказана.

7.2.2 Представление регулярного события конечным автоматом

Рассмотрим один способ задания регулярных событий с помощью графов.

Обобщенным источником в алфавите A называется ориентированный граф, с выделенными начальной и конечной вершинами v и w, дугам которого приписаны символы алфавита или пустые слова Λ. Допускается наличие петель и параллельных ребер. Последовательность символов, приписанных дугам некоторого пути $\pi(v_i, v_j)$ из одной вершины v_i в ту же или другую его вершину v_j, после исключения из неё символов λ образует некоторое слово в алфавите A, если в этой последовательности имеется хотя бы один символ из алфавита A.

Обозначим $\theta(u, \alpha)$ множество вершин u', таких, что существует путь $\pi(u, u')$ из u в u', которому соответствует слово α.

Каждый обобщенный источник G с начальной вершиной v и финальной вершиной w определяет событие $|G|$, образованное непустыми словами, соответствующими всевозможным путям $\pi(v,w)$ из начальной вершины в конечную вершину:

$$|G| = \{\alpha | \alpha \in A^*, \alpha \neq \Lambda, w \in \theta(v,\alpha)\}.$$

Лемма 4.4. *Если событие R регулярно, то существует обобщенный источник G, такой, что $R = |G|$.*

Доказательство осуществляется индукцией по шагам построения регулярного события R с помощью операций объединения, произведения и итерации, начиная от пустого или одноэлементных событий (в базисе индукции).

Лемма 4.5. *Если G — обобщенный источник, то событие $|G|$ представимо.*

Доказательство. Пусть $M = \{q_1, \dots q_n\}$ — множество вершин обобщенного источника G, в алфавите A, v, w — его начальная и финальная вершины. Обозначим

$Q = P(M)$ — множество всех подмножеств множества M.

Определим инициальный конечный автомат с множеством состояний Q и начальным состоянием $q_1 = \{v\}$.

$$V_{q_1} = (A, Q, \{0,1\}, \varphi, \psi, \{v\}),$$

$$\varphi(q,a) = \bigcup_{v' \in q} \theta(v', a),$$

$$\psi(q,a) = \left\{ \begin{array}{c} 1, \text{ если } w \in \varphi(q,a), \\ 0, \text{ если } w \notin \varphi(q,a) \end{array} \right. .$$

Индукцией по длине слова можно показать, что для любого непустого слова $\alpha, \alpha \in A^*$, выполняется $\varphi(\{v\}, \alpha) = \theta(v, \alpha)$.

Базис индукции:

$$\varphi(\{v\}, a) = \bigcup_{v' \in \{v\}} \theta(v', a) = \theta(v, a).$$

Предположим, что для любого непустого слова $\alpha, \alpha \in A^*$, выполняется $\varphi(\{v\}, \alpha) = \theta(v, \alpha)$.

Тогда (индуктивный переход)

$$\varphi(\{v\}, \alpha a) = \bigcup_{v' \in \varphi(\{v\}, \alpha)} \theta(v', a) = \bigcup_{v' \in \theta(v, \alpha)} \theta(v', a) = \theta(v, \alpha a).$$

Отсюда

$$\psi(v,\alpha) = \left\{ \begin{array}{c} 1, \text{ если } w \in \varphi(\{v\},\alpha) = \theta(v,\alpha), \\ 0, \text{ если } w \notin \varphi(\{v\},\alpha) = \theta(v,\alpha). \end{array} \right. .$$

Если $\alpha \in |G|$, то $w \in \theta(v, \alpha) = \varphi(\{v\}, \alpha)$ и по определению функции ψ имеем $\psi(\{v\}, \alpha) = 1$;

Если же $\alpha \notin |G|$, то $w \notin \theta(v, \alpha)$ и $\psi(\{v\}, \alpha) = 0$.

Таким образом, событие $|G|$ представимо в автомате V_{q_1} с помощью множества $\{0, 1\}$. Лемма доказана.

7.2.3 Теорема Клини

Как следствие лемм 4.3, 4.4, 4.5 получаем следующее утверждение

Теорема 4.1 (Клини). *Событие E в алфавите A представимо тогда и только тогда, когда оно регулярно.*

Множество представимых событий в конечном алфавите счетно; в то же время множество всех слов в нем имеет мощность континуум.

Приведем пример события, не являющегося представимым и следовательно регулярным.

Таким событием является множество всех слов в алфавите $\{0, 1\}$, имеющих одинаковое количество нулей и единиц. Предположим, что оно представимо автоматом $V_q = (\{0, 1\}, Q, B, \varphi, \psi, q)$ с помощью $B', B' \subseteq B$. В силу конечности множества Q найдутся такие $i_1,\ i_2,\ i_1 \neq i_2$,, что $\varphi(q, 0^{i_1}) = \varphi(q, 0^{i_2})$, но тогда

$$\psi(q_1, 0^{i_1} 1^{i_1}) = \psi(q_1, 0^{i_2} 1^{i_1}),$$

что невозможно, так как $\psi(q_1, 0^{i_1} 1^{i_1}) \in B'$, а $\psi(q_1, 0^{i_2} 1^{i_1}) \notin B'$, поскольку $i_1 \neq i_2$.

7.3 О проблеме равенства регулярных событий

Возникает вопрос о существовании алгоритма проверки равенства или неравенства двух регулярных событий R и R', заданных регулярными выражениями $\mathcal{R}$ и $\mathcal{R}'$. Эта проблема равенства имеет решение и сводится к проблеме неотличимости состояний двух автоматов следующим образом.

По двум заданным регулярным выражениям $\mathcal{R}$ и $\mathcal{R}'$ построим обобщённые источники G и G',, определяющие события $|G| = R$ и $|G'| = R'$, описываемые этими регулярными выражениями. Затем построим два инициальных конечных автомата,

$$V_q = (A, Q, \{0, 1\}, \varphi, \psi, q)$$

и

$$V'_{q'} = (A, Q', \{0, 1\}, \varphi', \psi', q')$$

представляющие те же события с помощью множества $\{1\}$.

Нетрудно видеть, что $R = R'$ тогда и только тогда, когда состояния q автомата V_q и q' автомата $V'_{q'}$ неотличимы.

Как было показано в лекции 3 (Теорема 3.3) множество $A^{|Q_1|+|Q_2|-1}$, отличает указанные состояния, если они отличимы.

Таким образом, проблема равенства регулярных событий разрешима.

7.4 Упражнение

7.1. Постройте автомат, представляющий регулярное множество

$$11(00^* \vee (11 \vee 01)^*)$$

Контрольные вопросы

1. Какие множества слов в данном алфавите называются регулярными событиями?

2. Как связаны между собой понятия регулярного выражения и регулярного события?

4. Какие регулярные выражения называются эквивалентными и в чем состоит проблема эквивалентности регулярных выражений?

5. На каких леммах основано доказательство регулярности события, представимого конечным автоматом?

6. Как построить обобщенный источник, определяющий событие, заданное регулярным выражением?

7. Как по обобщенному источнику построить автомат, представляющий определяемое им событие?

8. Как доказать или опровергнуть эквивалентность двух регулярных выражений?

Литература.

В.Б.Кудрявцев, С.В.Алёшин, А.С.Подколзин. Введение в теорию автоматов. М.:Наука, 1985.