КИБЕРНЕТИКА

А.А. ГРИГОРЯН, С.К. ШУКУРЯН

УДК 519.681

ЭКВИВАЛЕНТНОСТЬ ДВУМЕРНЫХ МНОГОЛЕНТОЧНЫХ АВТОМАТОВ

Ключевые слова: *автоматы, многоленточный автомат, многомерный автомат, эквивалентность.*

Автоматы с многомерными лентами, в которых движение головок монотонно во всех направлениях (движение назад невозможно), введены в [1]. Было показано, что проблема эквивалентности в классе схем программ над невырожденным базисом ранга единица сводится к эквивалентности многомерных многоленточных автоматов.

В настоящей статье рассматривается проблема эквивалентности двумерных многоленточных автоматов. Доказывается, что данная проблема сводится к эквивалентности многоленточных автоматов [2].

Приведем некоторые определения из [1], необходимые для дальнейшего рассмотрения.

Пусть $r>0$ — целое число, $N=\{0,1,\ldots\}$. Множество N^r называется r-мерной лентой, элементы множества N^r — r-ки вида $(a_1,\ldots,a_r)$ — называются ячейками ленты, а числа $a_1,\ldots,a_r$ — координатами соответствующей ячейки. Ячейка $(0,\ldots,0)$ является начальной. Пусть X — конечный алфавит, тогда отображение $N^r \to X$ называется заполнением r-мерной ленты символами из X.

Множество $S=\{(n_1,m_1),\ldots,(n_k,m_k)\}$, где n_i, m_i $(1\le i\le k)$ — натуральные числа и $n_i=n_j \Leftrightarrow i=j$ для любых $1\le i,j\le k$, называется сигнатурой многомерного многоленточного автомата. Сигнатура определяет количество и арность лент: если $(n,m)\in S$, то автомат с сигнатурой S имеет в точности m лент размерности n.

Рассмотрим случай, когда $n_i\le 2$. Будем считать, что $S=\{(1,m_1),(2,m_2)\}$, где $m_1\ge 0$, $m_2\ge 0$, $m=m_1+m_2>0$.

Пусть $A=\langle Q=Q_1\cup\ldots\cup Q_m, X, q_0, Q_F, \varphi, \psi\rangle$ — двумерный m-ленточный автомат с сигнатурой S, где Q — множество состояний, Q_i содержит все те и только те состояния, в которых считывается лента i, $Q_i\cap Q_j=\varnothing$, если $i\ne j$, X — входной алфавит, q_0 — начальное состояние, Q_F — множество заключительных состояний, $\varphi: Q\times X\to Q$ — функция переходов, $\psi: Q\times X\to\{1,2\}$ — функция направления движения.

Заполненную часть r-мерной ленты, в которой сумма координат каждой ячейки меньше или равна $n-1$, назовем r-мерным словом длины n. Множество всех r-мерных слов над алфавитом X обозначим $\Omega_r(X)$. r-мерное слово длины n распознается автоматом, если он при работе над словом попадает в заключительное состояние после чтения из ячейки, сумма координат которой равна $n-1$.

Заполненную часть r-мерной ленты, в которой сумма координат каждой ячейки равна k, назовем диагональю k слова и обозначим d_k. Первые (последние) $i\,(0< i \le k)$ символов диагонали d_k обозначим $d_k(i\_)\,(d_k(\_i))$. Длина слова совпадает с числом его диагоналей.

Пусть дана сигнатура $S=\{(1, m_1), (2, m_2)\}$, где $m_1 \ge 0$, $m_2 \ge 0$, $m = m_1 + m_2 > 0$. Набор одномерных и двумерных слов $(p_1, \ldots, p_m)$ длины m назовем m-ленточным словом с сигнатурой S, если число одномерных слов равно m_1 и число двумерных слов равно m_2. Без ограничения общности будем считать, что m-ленточные слова сигнатуры S имеют вид $(p_1, \ldots, p_{m_1},\ p_1', \ldots, p_{m_2}')$, где $p_1, \ldots, p_{m_1}$ — одномерные слова, $p'_1, \ldots, p'_{m_2}$ — двумерные слова. Если $m_2=0$, т.е. все ленты одномерные, то для таких m-ленточных слов в дальнейшем сигнатура указываться не будет.

Если автомат A распознает/не распознает слово w, то обозначим это $A(w)=1\,/\,A(w)=0$ соответственно.

Двумерные многоленточные автоматы A_1 и A_2 назовем эквивалентными и обозначим $A_1 \sim A_2$, если для каждого слова w $A_1(w)=A_2(w)$, причем если $A_1(w)=A_2(w)=1$, то позиции (координаты) головок на двумерных лентах совпадают.

Для того чтобы определить эквивалентность по множеству распознаваемых слов, для каждой двумерной ленты добавляется одна одномерная лента с алфавитом $\{1\}$. Автомат модифицируется так, чтобы считывать «1» с данной ленты каждый раз при движении в первом направлении на соответствующей двумерной ленте. Таким образом, длина слова на дополнительной ленте определяет позицию головки на текущей диагонали двумерной ленты и два автомата эквивалентны, если они распознают одинаковое множество слов.

Далее будем полагать, что двумерные автоматы уже содержат эти дополнительные ленты, и не будем упоминать их явно (они считаются учтенными в m_1).

Покажем, как моделировать процесс вычисления двумерного автомата A с сигнатурой S с помощью (m_1+2m_2)-ленточного автомата A' с одномерными лентами.

Пусть $\Omega=\Omega_1(X)\times\Omega_1(\{1,*\})$ — множество 2-ленточных слов с алфавитом первой ленты X и алфавитом второй ленты $\{1,*\}$. Определим отображение $\delta:\Omega_2(X)\to\Omega$.

Пусть дано двумерное слово w длины n. Слово $\delta(w)$ строится следующим образом (рис 1). Содержимое этого двумерного слова представляется на первой ленте $\delta(w)$ в следующем порядке: $x_{00}x_{01}x_{10}x_{02}x_{11}x_{20}\ldots x_{0,n-1}\ldots x_{n-1,0}$ (диагональ 0, диагональ 1,..., диагональ $n-1$). Часть заполнения ленты, соответствующую диагонали k двумерного слова (ячейки от $k(k+1)/2$ до $k(k+1)/2+k$ включительно), также назовем диагональю k и обозначим d_k.

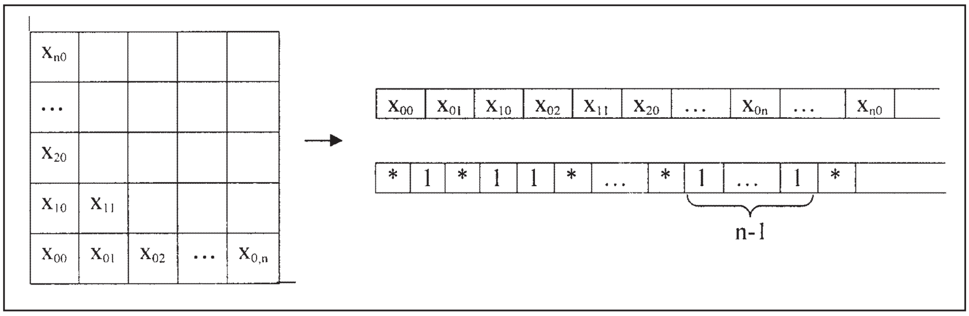

Рис. 1

Вторая лента $\delta(w)$ содержит длины этих диагоналей, уменьшенные на 1: $0, 1, 2, \ldots, n-2$. Они представляются в форме последовательности символов «1», отделенных один от другого символом «*» (например, *1*11* ...*1...1*). Пусть ω — конечное 2-ленточное слово из Ω и его вторая лента содержит по меньшей мере $k+1$ символов типа «*». Число символов типа «1» между k и $k+1$ символами типа «*» обозначим $T_k(\omega)$ ($T_0(\omega)$ — число символов типа «1» перед первым символом типа «*»). В приведенном случае $T_k(\delta(w)) = k$ для $0 \le k \le n-2$.

Пусть $W = \Omega_1(X)^{m_1} \times \Omega_2(X)^{m_2}$ — множество всех m-ленточных слов сигнатуры S. Обозначим $W_1 = \Omega_1(X)^{m_1} \times \Omega^{m_2}$.

Отображение δ расширяется на множество всех m-ленточных слов сигнатуры S следующим образом: $\delta: W \to W_1$ для m-ленточного слова $w = (p_1, \ldots, p_{m_1}, p'_1, \ldots, p'_{m_2}) \in W$, $\delta(w) = (p_1, \ldots, p_{m_1}, p_1^{(1)}, p_1^{(2)}, \ldots, p_{m_2^{(1)}}, p_{m_2^{(2)}})$, где $(p_i^{(1)}, p_i^{(2)}) = \delta(p'_i)$ для $i = 1, \ldots, m_2$.

Для слова w сигнатуры S $\delta(w)$ является $(m_1 + 2m_2)$-ленточным словом сигнатуры $\{(1, m_1 + 2m_2)\}$.

Первая лента содержит последовательность диагоналей, вторая — расстояния между символами, расположенными в соседних ячейках на двумерной ленте. Для моделирования с помощью одномерных лент движения из одной ячейки в другую на двумерной ленте необходимо «пропустить»/проигнорировать первые l символов на первой одномерной ленте, где l — длина текущей диагонали в случае, когда моделируется движение в первом направлении на двумерной ленте, и проигнорировать первые $l+1$ символов в случае, когда моделируется движение во втором направлении. В то же время это приводит к множеству слов на второй одномерной ленте, которое не является регулярным, т.е. невозможно построить конечный автомат, распознающий рассмотренное множество 2-ленточных слов.

Чтобы преодолеть возникшие трудности, расширим рассматриваемое множество 2-ленточных слов с помощью следующих правил.

1. Для данной диагонали двумерной ленты считаем кодом диагонали любую последовательность, которая помимо исходных символов диагонали, записанных в том же порядке, содержит некоторые избыточные символы после исходных, при этом вместо длины исходной диагонали на второй ленте должна быть записана длина «расширенной» диагонали.

2. Если у данной пары соседних диагоналей существует последовательность символов, которая заканчивает первую диагональ и одновременно начинает вторую, и если эта последовательность опущена в одной из диагоналей во время их кодирования на первой одномерной ленте, то полученная «усеченная» пара диагоналей также считается кодом двух соседних диагоналей. При этом вместо длины исходной диагонали на второй ленте должна быть закодирована длина «усеченной» диагонали.

Покажем, что такое расширение приводит к регулярному множеству слов на второй ленте за счет того, что больше нет необходимости иметь возрастающие на единицу длины диагоналей. Покажем также, что существует очевидное отображение этого расширенного множества на исходное множество 2-ленточных слов, т.е. по заданной паре слов на простых лентах, распознаваемых 2-ленточным автоматом, можно однозначно получить заполнение соответствующей двумерной ленты, распознаваемой исходным автоматом. Для множества всех двумерных слов результирующим множеством слов на второй ленте будет $\{\{1\}^*\}$.

Рассмотрим следующее подмножество множества $W_1 : W' = \{w' \mid \exists w \in W,$ $w' = \delta(w)\}$, т.е. слова, у которых на второй ленте написано $0, 1, 2, \ldots$. Легко убедиться, что верна следующая лемма.

Лемма 1. Отображение $\delta : W \to W'$ взаимно однозначно.

Построим автомат A', который работает над словами из множества W_1. Главная трудность его построения — реализация описанных выше правил, которые позволяют преобразовать исходное нерегулярное множество слов второй ленты в регулярное. Леммы 2 и 3 определяют шаги преобразования исходного множества в расширенное регулярное множество с сохранением некоторых базовых свойств исходного множества.

Множество состояний A' обозначим Q'. Для каждого состояния $q \in Q$ в A' определяется соответствующее состояние q' такое, что $q' \in Q_F' \Leftrightarrow q \in Q_F$ и $q' \in Q'_k \Leftrightarrow q \in Q_k$. Для каждого состояния $q_1 \in Q_i$, где лента i двумерна, и входного символа x A' будет иметь два дополнительных состояния, которые используются для передвижения головки автомата A' на первой одномерной ленте при моделировании движения головки на двумерной ленте автомата A. Фрагмент автомата A', соответствующий одному переходу в A, показан на рис. 2 для случаев $\psi(q_1, x) = 1$ и $\psi(q_1, x) = 2$ (заштрихованные состояния соответствуют второй одномерной ленте). Если лента i — одномерна, то соответствующий фрагмент графа переходов автомата A повторяется в A'.

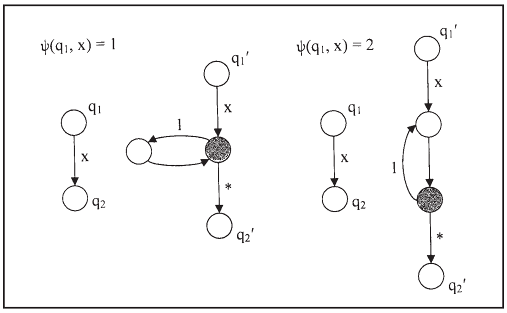

Рис. 2

Далее $(m_1 + 2m_2)$-ленточный автомат A', построенный указанным способом из исходного двумерного m-ленточного автомата с сигнатурой S, обозначим $\gamma(A)$.

При моделировании чтения символа x на двумерной ленте в состоянии q автомата A автомат A' работает следующим образом. После чтения символа x на первой одномерной ленте в зависимости от значения $\psi(q, x)$ пропускаются l или $l+1$ символов (l — число символов «1» на второй одномерной ленте), затем происходит переход в состояние, соответствующее $\varphi(q, x)$. Пропуск символов означает, что поочередно считываются символы со второй и первой лент до тех пор, пока со второй ленты не будет считан символ «*».

Лемма 2. Пусть $w \in W$ — m-ленточное двумерное слово с сигнатурой S, A — двумерный m-ленточный автомат с сигнатурой S, $A' = \gamma(A)$. Тогда $A'(\delta(w)) = A(w)$.

Доказательство. Покажем, что после произвольного шага k автомата A, где $k = k_1 + \ldots + k_m$, k_i — число шагов на ленте i, и соответствующей последовательности шагов автомата A', моделирующих шаг автомата A:

1) A и A' перейдут в состояния, соответствующие описанным построениям;

2) текущий символ слова w, обозреваемый активной головкой автомата A, и текущий символ на соответствующей одномерной ленте автомата A', обозреваемый его активной головкой, совпадут;

3) текущий символ на одномерной ленте с кодом длины диагонали, на которой находится активная головка автомата A', является первым символом «1» в соответствующем коде.

Доказательство проводится индукцией по k. Рассмотрим только двумерные ленты, так как поведение A' для одномерных лент не отличается от поведения автомата A.

Поскольку автомат A' полностью повторяет движения автомата A на одномерных лентах, без ограничения общности можно считать, что начальное состояние A $q_0 \in Q_l$, где лента l — двумерна. Как описано выше, двумерная лента l моделируется с помощью трех одномерных лент, называемых далее первой, второй и третьей одномерными лентами l соответственно.

Если $k=1$, то A перейдет в состояние $a=\varphi(q_0, x_{00})$, A' — в состояние a'. Текущим символом слова w на двумерной ленте l и текущим символом $\delta(w)$ на первой одномерной ленте l будут или $x_{01}^{(l)}$, или $x_{10}^{(l)}$ (диагональ 1), головка на второй одномерной ленте l будет обозревать единственный символ «1» кода длины 1.

Предположим, что утверждение верно для первых $k-1$ шагов. Это означает, что:

1) автомат A находится на диагонали k_i в состоянии $q \in Q_i$;

2) автомат A' находится в соответствующем состоянии $q' \in Q_i'$;

3) автомат A' находится на диагонали k_i на первой одномерной ленте i и в начале кода длины диагонали k_i на второй одномерной ленте i.

Пусть x — текущий символ на ленте i слова w и первой одномерной ленте i слова $\delta(w)$. Тогда согласно построению автомат A' продвинется на первой ленте i столько раз, сколько символов типа «1» кода длины находится справа от головки ленты до первого символа «*». По предположению индукции головка второй ленты i обозревает первый символ кода диагонали k_i, поэтому A' продвинется k_i раз на первой ленте i. Он продвинется еще раз, если $\psi(q, x)=2$. Головка A' на первой ленте i будет обозревать символ диагонали k_i+1. При этом текущий символ на второй ленте i автомата A' станет первым символом «1» кода диагонали k_i+1. После этого A' попадает в состояние $a' \in Q'_j$, соответствующее состоянию $a=\varphi(q, x) \in Q_j$ автомата A. Поскольку положение головок для первой, второй и третьей лент j автомата A' не изменилось в рассмотренном выше процессе, то предположение индукции для них остается в силе. Итак, предположение шага индукции верно и для k, т.е. поскольку $a' \in Q_F' \Leftrightarrow a \in Q_F$, то $A'(\delta(w)) = A(w)$.

Лемма доказана.

Лемма 3. Для любого (m_1+2m_2)-ленточного слова $w_1=(p_1, \ldots, p_{m_1},\ p_1^{(1)}, p_1^{(2)}, \ldots, p_{m_2^{(1)}}, p_{m_2^{(2)}}) \in W_1$ существует слово $w \in W$ такое, что для любого двумерного m-ленточного автомата A с сигнатурой S $\quad A'=\gamma(A) \to \to A'(w_1) = A'(\delta(w))$.

Доказательство. Если $\exists\ w \in W$ такое, что $w_1=\delta(w)$, то очевидно, что лемма верна. Предположим, что это не так.

Пусть k — наименьшее число, для которого $T_k(p_i^{(1)}, p_i^{(2)}) \neq k$. Обозначим $s=T_k(p_i^{(1)}, p_i^{(2)})$. Возможны два случая.

1. **Случай** $s>k$. Тогда $s-k$ символов после диагонали k игнори- руются автоматом A'. Если их удалить и соответственно удалить $s-k$ символов типа «1» на второй ленте, то согласно построению поведение автомата A' на слове w_1 не будет отличаться от его поведения на результирующем слове (рис. 3).

2. **Случай** $s<k$. Последние $k-s$ символов являются недостающей частью одной из двух соседних диагоналей для автомата A'. Поэтому если их повторить

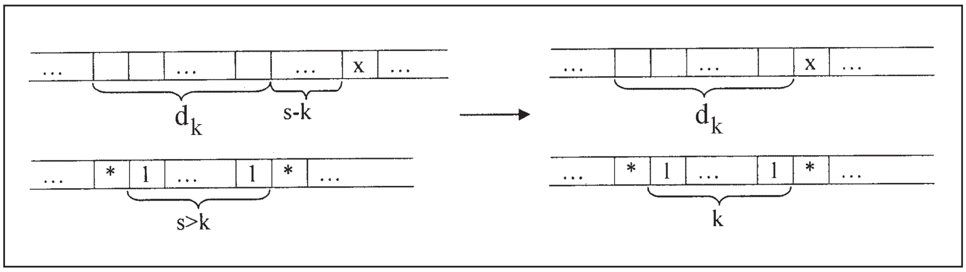

Рис. 3

и соответственно добавить $k-s$ символов типа «1» в код длины диагонали на второй ленте, то согласно построению поведение автомата A' на слове w_1 не будет отличаться от его поведения на результирующем слове (рис. 4).

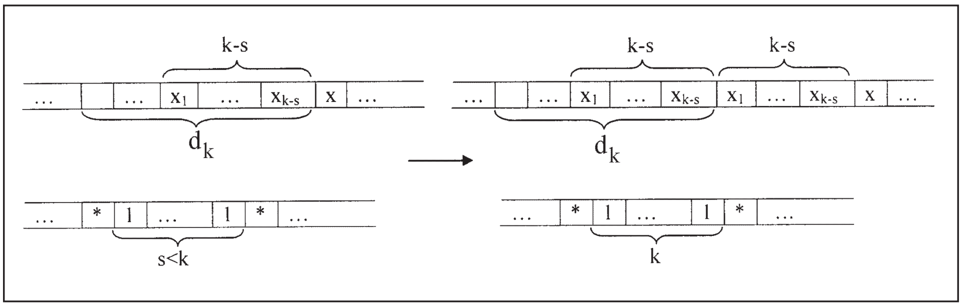

Рис. 4

Считая, что данные преобразования выполнены для всех $i=1,\ldots,m_2$, получим $w'=(p_1,\ldots,p_{m_1},u_1^{(1)},u_1^{(2)},\ldots,u_{m_2^{(1)}},u_{m_2^{(2)}})\in W'$. Таким образом, $w=\delta^{-1}(w')$ и $A'(w_1)=A'(\delta(w))$.

Лемма доказана.

Следствие 1. Существует факторизация множества W_1 такая, что каждый класс содержит одно и только одно слово из W' и для произвольного автомата A с сигнатурой $S\ \gamma(A)$ распознает/не распознает все слова данного класса.

Например, слова $(x_{00}\{x\}^r x_{01}x_{10}\{x\}^s x_{02}x_{11}x_{20},\ \{1\}^r *1\{1\}^s *)$ и $(x_{00}x_{01}x_{10}x_{02}x_{11}x_{20},\ *1*)\in W'$ находятся в одном классе.

На рис. 5 имеем $A'(w')=A'(\delta(w))$ — последовательность «xx» будет пропущена автоматом, так как $T_1(w')=3$, а «$x_{11}x_{20}$» содержится в двух соседних диагоналях, поскольку $T_2(w')=0$.

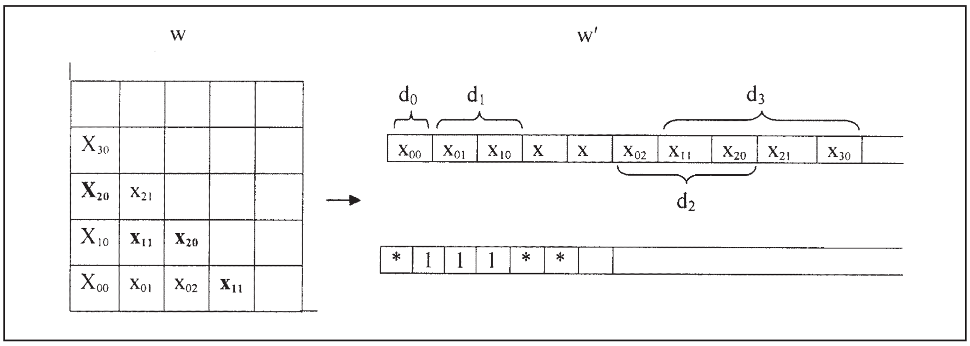

Рис. 5

Рассмотрим другой пример. Пусть дан 1-ленточный двумерный автомат A, распознающий только слова, у которых $x_{0i} = a, i = 0, 1, \ldots$ (рис. 6).

Автомат A' распознает слова, которые содержат символы типа «a» на первой ленте, отделенные один от другого последовательностью произвольных символов из X. При этом вторая лента содержит только длины этих последовательностей, закодированные с помощью символов «1» и «*». Пара $(axxxaxa, 111{*}1{*})$ является примером такого слова. В этом случае проекции распознаваемых 2-ленточных слов — множества $\{a\{x\}\}$ и $\{\{1\}{*}\}$ на первой и второй лентах соответственно.

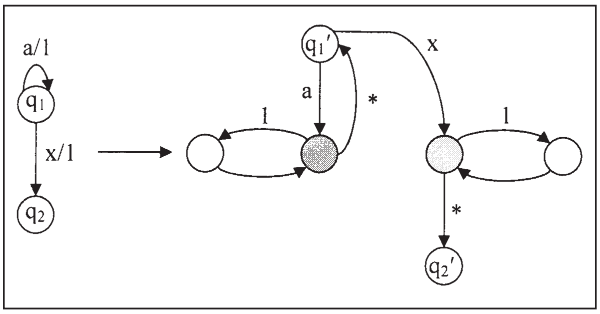

Рис. 6

Пусть даны два двумерных m-ленточных автомата A_1 и A_2 с сигнатурой S. Пусть $A_1' = \gamma(A_1)$ и $A_2' = \gamma(A_2)$.

Лемма 4.Справедливо следующее утверждение: $A_1' \sim A_2' \to A_1 \sim A_2$.

Доказательство.Предположим обратное: $A_1' \sim A_2'$, но A_1 и A_2 не эквивалентны. Это означает, что $\exists w, A_1(w) \neq A_2(w)$. Из леммы 2 следует, что $A_1(w) = A_1'(\delta(w))$ и $A_2(w) = A_2'(\delta(w))$. Отсюда получаем, что $A_1'(\delta(w)) \neq$ $\neq A_2'(\delta(w))$. Но это противоречит $A'_1 \sim A'_2$. Таким образом, предположение, что A_1 и A_2 не эквивалентны, неверно.

Лемма доказана.

Лемма 5. Справедливо следующее утверждение: $A_1' \not\sim A_2' \Rightarrow A_1 \not\sim A_2$.

Доказательство. Из $A_1' \not\sim A_2'$ следует, что $\exists w', A_1'(w') \neq A_2'(w')$. Если $w' \approx W'$, то согласно лемме 3 существует $w'' \in W', A_1'(w') = A_1'(w'')$ и $A_2'(w') = A_2'(w'')$, а значит, $A_1'(w'') \neq A_2'(w'')$. Таким образом, можно считать, что $w' \in W'$.

Из леммы 2 вытекает, что $A_1(\delta^{-1}(w')) = A_1'(w')$ и $A_2(\delta^{-1}(w')) = A_2'(w')$. Следовательно, $A_1(\delta^{-1}(w')) \neq A_2(\delta^{-1}(w'))$. Это означает, что A_1 и A_2 не эквивалентны.

Лемма доказана.

Леммы 4 и 5 приводят к следующей теореме.

Теорема. Справедливо утверждение $A_1 \sim A_2 \Leftrightarrow A_1' \sim A_2'$.

Следствие 2. Проблема эквивалентности двумерных многоленточных автоматов разрешима.

Это следует из теоремы и того факта, что проблема эквивалентности многоленточных автоматов разрешима [2].

Авторы выражают глубокую благодарность члену-корреспонденту НАН Украины А.А. Летичевскому, определившему более 30 лет назад общее направление исследований, профессору Д. Кнуту, существенно повлиявшему на возобновление исследований, и доктору физ.-мат. наук А.Б. Годлевскому за постоянный интерес к работе, ценные советы и замечания.

СПИСОК ЛИТЕРАТУРЫ

1. Годлевский А.Б., Летичевский А.А., Шукурян С.К. О сводимости

проблемы функциональной эквивалентности схем программ над невырожденным базисом ранга единица к эквивалентности автоматов с многомерными лентами // Кибернетика. — 1980. — № 6. — С. 1–7.

2. Harju T., Karhumaki J. The equivalence problem of multitape finite automata // Theoret. Comput. Sci. — 1991. — **78**. — P. 347–355.

Поступила 24.04.2007