О преобразовании и автоматном представлении ε-нечетко-автоматных грамматик

А. В. Светлаков

***Аннотация*—Работа посвящена анализу ε-нечетко-автоматных грамматик. Доказана теорема об удалении правила вида $B \to \varepsilon\,(n)$ из таких грамматик с приведением алгоритма, также доказана теорема об удалении аксиомы из правых частей таких грамматик, аналогично с приведением алгоритма.**

Помимо этого, в статье рассмотрены грамматики, содержащие правило $S \to \varepsilon\,(n)$, и способ их автоматного представления. Введены ф-НДКА (нечеткие детерминированные конечные автоматы с фантомным состоянием) и НДКА с предопределением (с предопределенной степенью принадлежности начального состояния). В виде теорем сформулирована связь между грамматиками и этими автоматами, и приведены алгоритмы преобразований.

***Ключевые слова*—нечетко-автоматная грамматика, ε-нечетко-автоматная грамматика, нечеткие детерминированные конечные автоматы с фантомным состоянием, нечеткие детерминированные конечные автоматы с предопределением, нечетко-автоматный язык, фантомное состояние.**

I. Введение

Математическая лингвистика включает в себя исследование формальных языков различных типов, которые на сегодняшний день имеют множество различных приложений – от компиляторов до переводчиков [6]. Ее можно расширить, воспользовавшись результатами теории нечетких множеств Заде [4], получив, таким образом, нечеткие языки.

Нечеткие языки в математической лингвистике представляют собой такой класс языков, в которых каждое слово принадлежит ему с некоторой степенью принадлежности, заданной функцией $\mu_L(x): T^* \to [0,1]$ [1]. Такой язык можно задать нечеткой грамматикой $\tilde{G} = \langle N, T, S, P, w, \otimes, \oplus \rangle$, где $\langle N, T, S, P \rangle$ – обычная грамматика Хомского. Функционал $w: P \to R$ ставит в соответствие каждому правилу из P его весовой коэффициент из множества действительных чисел R, $\otimes$ – левоассоциативная бинарная операция на $[0,1]^2 \to [0,1]$ или $R^2 \to [0,1]$ (это зависит от весов), $\oplus$ – ассоциативная и коммутативная бинарная операция на $[0,1]^2 \to [0,1]$. Отметим, что в общем случае алгебра $\langle R, \otimes \rangle$ является магмой (группоидом) с естественными ограничениями на бинарную операцию. Алгебра $\langle [0,1], \oplus \rangle$ является коммутативной полугруппой.

Статья получена 27 сентября 2025.

Нечетким языком $\widetilde{L(G)}$, порожденным грамматикой $\tilde{G}$, называется множество $\widetilde{L(G)} = \{(x, \mu_L(x)) | S \Rightarrow^* x \in T^*, \mu_L(x) = \otimes_i w(p_i)\}$ при условии однозначности грамматики Хомского [3].

В случае неоднозначности грамматики Хомского принадлежность слова языку рассчитывается по формуле $\mu_L(x) = \oplus_j \left(\otimes_{ij} w(p_{ij})\right) = \oplus_j \left(\mu_j(x)\right)$ для всех j [3].

В общем случае такое определение включает в себя самые разнообразные нечеткие грамматики, которые можно классифицировать по их сложности: в частности, нечеткая автоматная грамматики и контекстно-свободная грамматика [2].

Определение. Грамматика называется нечетко-автоматной, если все ее правила имеют вид $A \to a$ или $A \to aB$ с весами, присвоенными по правилу w, где $A, B \in N$ и $a \in T$. Допускается наличие правила $S \to \varepsilon$, если $\mu_L(\varepsilon) > 0$, при условии, что S не встречается в правых частях правил.

Определение. Грамматика называется ε-нечетко-автоматной, если она является нечетко-автоматной, а также в ней присутствуют правила вида $A \to \varepsilon\,(n)$, или правило $S \to \varepsilon\,(n)$, когда S встречается в правых частях правил.

Наличие ε-правил в автоматной грамматике может приводить к тому, что грамматика оказывается неоднозначной: одно и то же слово может выводиться с разной степенью принадлежности. Назовем такие грамматики ε-неоднозначными, а ε-однозначными грамматиками будем называть такие ε-нечетко-автоматные грамматики, у которых наличие ε-правила не приводит к неоднозначности. Так образом, однозначная ε-нечетко-автоматная грамматика может быть только ε-однозначна, а неоднозначная может быть как ε-однозначной, так и ε-неоднозначной.

Пример. Пусть $S \to aA\,(0.5)$, $A \to a | aA\,(0.5)$, $A \to \varepsilon\,(2)$, причем $x \otimes y = x \cdot y$ и $x \oplus y = max(x, y)$. Очевидно, что, к примеру, слово a^3 имеет два вывода: $S \to aA \to aaA \to aaa$ и $\mu_1(a^3) = 0.125$, а также $S \to aA \to aaA \to aaaA \to aaa$ и $\mu_2(a^3) = 0.25$. Применяя вторую бинарную операцию $0.125 \oplus 0.25 = 0.25$, таким образом, степень принадлежности каждого слова определяется однозначно, в частности $\mu_L(a^3) =$

0.25. Нетрудно видеть, что данная грамматика эквивалентна следующей однозначной грамматике без ε-правил: $S \to a\,(1)$, $S \to aA\,(0.5)$, $A \to aA(0.5)$, $A \to a(1)$, потому что обе грамматики порождают один и тот же язык: $\tilde{L} = \left\{(a,1),(a^2,0.5),\ldots,\left(a^i,\mu_i = \frac{\mu_{i-1}}{2}\right),\ldots\right\}$.

Возникает следующий вопрос – всегда ли можно ε-нечетко-автоматную грамматику преобразовать в нечетко-автоматную автоматную грамматику?

II. Теорема об удалении правила $B \to \varepsilon\,(n)$

Рассмотрим произвольную ε-нечетко-автоматную грамматику без правил вида $S \to \varepsilon\,(n)$. В такой грамматике всегда будет присутствовать правило вида $A \to \varepsilon\,(n)$. Кроме того, в выводе любого слова такое правило всегда будет применяться последним, таким образом, оно повлияет лишь на принадлежность слов с этим примененным правилом.

Лемма. Из грамматики, содержащей правила $S \to aA\,(m)$, $A \to \varepsilon\,(n)$, и не содержащей таких правил, в которых нетерминал A встречается в правых частях правил, правило $A \to \varepsilon\,(n)$ всегда может быть удалено вне зависимости от вида бинарных операций.

Доказательство. Если грамматика ε-однозначна, рассмотрим вывод слова a в грамматике: $S \to aA\,(m) \to a\,(m \otimes n)$, и очевидно, что в грамматику всегда можно добавить правило $S \to a\,(m \otimes n)$, удалив $A \to \varepsilon\,(n)$.

Если грамматика ε-неоднозначна, то она обязательно содержит правило $S \to a(l)$ вывод слова a будет таким: $S \to aA\,(m) \to a\,(m \otimes n)$ или $S \to a(l)$ и $\mu(a) = l \oplus (m \otimes n)$. В таком случае можно добавить правило $S \to a\,(l \oplus (m \otimes n))$, удалив правила $A \to \varepsilon\,(n)$ и $S \to a(l)$.■

При рассмотрении следующих теорем, мы исключаем рассмотрение указанного в лемме случая.

Теорема (об удалении правила $B \to \varepsilon\,(n)$). 1) Если ε-нечетко-автоматная грамматика ε-однозначна и уравнение $a \otimes y = (a \otimes k) \otimes n$ имеет хотя бы одно решение относительно y в магме $\langle R,\otimes\rangle$, не зависимое от значения a, то существует нечетко-автоматная грамматика, эквивалентная данной.

2) Если ε-нечетко-автоматная грамматика ε-неоднозначна и уравнение $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$ имеет хотя бы одно решение относительно y в алгебре $\langle R,\otimes,\oplus\rangle$ не зависимое от значения a, то существует нечетко-автоматная грамматика, эквивалентная данной.

Алгоритм. Опишем алгоритм удаления правила вида $B \to \varepsilon\,(n)$ из грамматики.

1) Найти все ε-правила вида $B \to \varepsilon\,(n)$ в грамматике, B является ε-порождающим нетерминалом;

2) Найти все правила, содержащие в правых частях ε-порождающий нетерминал – это правила вида $A \to bB\,(k)$, а также все правила $A \to b\,(l)$.

3) Если грамматика ε-однозначна, решить уравнение $a \otimes y = (a \otimes k) \otimes n$ относительно y, где $a \in [0;1]$ – параметр. Если грамматика ε-неоднозначна, решить уравнение $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$ относительно y, где $a \in [0;1]$ – параметр. Выбрать то решение, которое не зависит от a.

4) Удалить все правила вида $A \to b\,(l)$ и $B \to \varepsilon\,(n)$, и добавить в грамматику правила вида $A \to b\,(y)$.

Доказательство. Покажем, что преобразованная по указанному алгоритму грамматика будет эквивалентна исходной грамматике.

Рассмотрим вывод слова α с $\mu(\alpha)$ в ε-нечетко-автоматной грамматике. Покажем, что такое же слово выводится преобразованной грамматике по алгоритму, указанному в теореме.

Рассмотрим вывод слова $\alpha b^{-1}A$ в ε-нечетко-автоматной грамматике. На всех этапах вывода применялись правила вида $N \to aC\,(s)$, которые не изменились после применения алгоритма, таким образом, слово $\alpha b^{-1}A$ выводится и в преобразованной грамматике с той же степенью принадлежности.

Рассмотрим последние шаги в выводе этого слова α. Если в грамматике существует правило $A \to b\,(l)$ и не существует правила $B \to \varepsilon\,(l)$, то $A \to b\,(l)$ существует и в преобразованной грамматике, и слово выводится за один шаг: $\alpha b^{-1}A \to \alpha$, причем $\mu(\alpha) = \mu(\alpha b^{-1}A) \otimes l$, как и в изначальной грамматике.

Если в грамматике нет правила $A \to b\,(l)$, тогда существуют правила $A \to bB\,(k)$ и $B \to \varepsilon\,(n)$, и слово α выводится за два шага из слова $\alpha b^{-1}A$: $\alpha b^{-1}A \to \alpha B \to \alpha$, причем $\mu(\alpha) = (\mu(\alpha b^{-1}B) \otimes k) \otimes n$. Пусть $\mu(\alpha b^{-1}A) = a$. В новой грамматике существует правило $A \to b\,(y)$ и слово выводится за один шаг: $\alpha b^{-1}A \to \alpha$, причем $\mu'(\alpha) = a \otimes y$. Так как y есть решение уравнения $a \otimes y = (a \otimes k) \otimes n$, то $\mu'(\alpha) = \mu(\alpha)$.

Наконец, если в исходной грамматике существует правило $A \to b\,(l)$, а также правила $A \to bB\,(k)$ и $B \to \varepsilon\,(n)$, то слово α выводится неоднозначно: $\alpha b^{-1}A \to \alpha$ и $\mu_1(\alpha) = a \otimes l$, а также $\alpha b^{-1}A \to \alpha B \to \alpha$ и $\mu_2(\alpha) = (a \otimes k) \otimes n$, и конечная степень принадлежности определяется так: $\mu(\alpha) = \mu_1(\alpha) \oplus \mu_2(\alpha) = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$. В новой грамматике существует правило $A \to b\,(y)$ и слово выводится за один шаг: $\alpha b^{-1}A \to \alpha$, причем $\mu'(\alpha) = a \otimes y$. Так как y есть решение уравнения $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$, то $\mu'(\alpha) = \mu(\alpha)$.

Рассмотрим вывод слова α с принадлежностью $\mu(\alpha) = q$ в преобразованной грамматике, и покажем, что такое же слово порождается исходной грамматикой. Последнее правило, которое применялось в этой грамматике – правило $A \to b\,(y)$, где y - решение уравнения $a \otimes y = (a \otimes k) \otimes n$ или $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$, согласно указанному алгоритму. Алгоритм меняет лишь последние правила, а значит сентенциальная форма $\alpha b^{-1}A$ и $\mu(\alpha b^{-1}A\,) = a$ существует в обоих грамматиках. Если исходная грамматика ε-однозначна, то, применяя алгоритм в обратном порядке, в исходной грамматике будут правила $A \to bB\,(k)$ и $B \to \varepsilon\,(n)$, и не будет правила $A \to b\,(y)$, тогда $\alpha b^{-1}A \to \alpha B \to \alpha$ и $\mu(\alpha) = (a \otimes k) \otimes n = q$. Если исходная грамматика ε-однозначна, то, применяя алгоритм в обратном порядке, в исходной грамматике будут правила $A \to bB\,(k)$, $B \to \varepsilon\,(n)$ и $A \to b\,(l)$, и слово выводится двумя способами: $\alpha b^{-1}A \to \alpha$ и $\mu_1(\alpha) = a \otimes l$, а также $\alpha b^{-1}A \to \alpha B \to \alpha$ и

$\mu_2(\alpha) = (a \otimes k) \otimes n$, причем $\mu(\alpha) = \mu_1(\alpha) \oplus \mu_2(\alpha) = q$. Теорема полностью доказана. ■

Замечание. В теореме имеются в виду грамматики, содержащие исключительно правила приводящие либо к ε-однозначности, либо к ε-неоднозначности. Если грамматики содержит и те, и другие правила, то два пункта теоремы должны выполняться одновременно.

Пример. Рассмотрим предыдущий пример. Пусть $S \to aA\ (0.5)$, $A \to a|aA\ (0.5)$, $A \to \varepsilon\ (2)$, причем $x \otimes y = x \cdot y$ и $x \oplus y = max(x, y)$. Как было показано, грамматика является ε-неоднозначной, то есть для того, чтобы можно было убрать ε-правило, нужно решить уравнение $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$, в данном случае: $a \cdot y = \max\big((a \cdot l), (a \cdot k \cdot n)\big)$. По крайней мере одним из решений уравнения является $y = \max(l, k \cdot n)$, которое не зависит от параметра a, а это значит, что ε-правило можно удалить. В данной грамматике $y = \max(0.5, 0.5 \cdot 2) = 1$, так что добавляем правила $A \to a\ (1)$ и $S \to a\ (1)$, удалив правила $A \to a\ (0.5)$ и $A \to \varepsilon\ (2)$.

III. Следствия из теоремы

Следствие 1. Если уравнение $a \otimes y = (a \otimes k) \otimes n$ в ε-однозначной ε-нечетко-автоматной грамматике или уравнение $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$ в ε-неоднозначной грамматике не имеют решений относительно y хотя бы для одного $a = \mu(\alpha b^{-1}A)$, где $\alpha b^{-1}A$ – это сентенциальная форма из которой выводится слово α максимум за 2 шага с использованием ε-правила, то не существует эквивалентной нечетко-автоматной грамматики, кроме случая, описанного в лемме.

Доказательство. Предположим, что такая грамматика существует. В случае ε-однозначности грамматики, если $\mu(\alpha b^{-1}A) = a$, $A \to bB\ (k)$ и $B \to \varepsilon\ (n)$, то $\mu(\alpha) = (a \otimes k) \otimes n$, с другой стороны, в предполагаемой нечетко-автоматной грамматике нет правила $B \to \varepsilon\ (n)$, поэтому должно быть правило $A \to b\ (y)$, то есть $\alpha b^{-1}A \to \alpha$, тогда $\mu'(\alpha) = a \otimes y$, а поскольку уравнение $a \otimes y = (a \otimes k) \otimes n$ не имеет решения, то $\mu'(\alpha) \neq \mu(\alpha)$ при любом y.

В случае ε-неоднозначности грамматики $\mu(\alpha b^{-1}A) = a$, $A \to b\ (l)$, a, $A \to bB\ (k)$ и $B \to \varepsilon\ (n)$, то есть конечная принадлежность $\mu(\alpha) = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$. С другой стороны, в предполагаемой нечетко-автоматной грамматике нет правила $B \to \varepsilon\ (n)$, поэтому должно быть единственное правило $A \to b\ (y)$, то есть $\alpha b^{-1}A \to \alpha$ и $\mu'(\alpha) = a \otimes y$, а поскольку уравнение $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$ не имеет решения, то $\mu'(\alpha) \neq \mu(\alpha)$ при любом y. ■

Пример. Рассмотрим грамматику с правилами $S \to aA\ (0.1)$, $A \to aA(0.2)$, $A \to \varepsilon\ (0.3)$ и $x \otimes y = \begin{cases} 0.2, \text{если } x = 0.1 \\ 0.3, \text{если } x = 0.2 \\ 0.4, \text{в прочих случаях} \end{cases}$. Слово a^2 в такой грамматике выводится так: $S \to aA \to a^2A \to a^2$ со степенью принадлежности $\mu(a^2) = 0.1 \otimes 0.2 \otimes 0.3 = 0.3$, в новой грамматике должно быть правило $A \to a(y)$, где $0.1 \otimes y = 0.3$, но такое уравнение не имеет решений. Стало быть, удалить $A \to \varepsilon\ (0.3)$ из грамматики нельзя.

Следствие 2. Если уравнение $a \otimes y = (a \otimes k) \otimes n$ в ε-однозначной ε-нечетко-автоматной грамматике или уравнение $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$ в ε-неоднозначной грамматике имеют решения, каждое из которых зависит от параметра a, то для этой грамматики существует эквивалентная нечетко-автоматная грамматика только в двух случаях: 1) слово вида $\alpha b^{-1}A = S$ и не равно ничему другому или $\alpha b^{-1}A$ имеет единственную степень принадлежности для любого α, выводимого в грамматике; 2) слово вида $\alpha b^{-1}A$ имеет такое подмножество степеней принадлежности, что хотя бы одно из значений решений указанных уравнений равны.

Доказательство. Случай $\forall\alpha\colon \alpha b^{-1}A = S$ доказан в лемме. Рассмотрим $\exists\alpha\colon \alpha b^{-1}A \neq S$.

Пусть грамматика ε-однозначна. Если значение $\mu(\alpha b^{-1}A) = a$ единственно, то значение y в одном из решений уравнения $a \otimes y = (a \otimes k) \otimes n$ будет единственным, то есть правило $A \to b\ (y)$ в эквивалентной грамматике будет единственным, таким образом, эквивалентная нечетко-автоматная грамматика существует согласно теореме выше.

Теперь пусть значение $\mu(\alpha b^{-1}A)$ при разных α будет неоднозначным, выберем любые два значения – $\mu(\alpha_1 b^{-1}A) = a_1$, $\mu(\alpha_2 b^{-1}A) = a_2$. В ε-однозначной грамматике есть правила $A \to bB\ (k)$ и $B \to \varepsilon\ (n)$, а значит, $\mu(\alpha_1) = (a_1 \otimes k) \otimes n$ и $\mu(\alpha_2) = (a_2 \otimes k) \otimes n$. В предполагаемой нечетко-автоматной грамматике нет правила $B \to \varepsilon\ (n)$, поэтому должно быть правило $A \to b\ (y)$, то есть $\alpha_1 b^{-1}A \to \alpha_1$ и $\alpha_2 b^{-1}A \to \alpha_2$. Здесь $\mu'(\alpha_1) = a_1 \otimes y$ и $\mu'(\alpha_2) = a_2 \otimes y$. Без потери общности, пусть y такое, что $\mu(\alpha_1) = \mu'(\alpha_1)$, то есть является решением уравнения $a_1 \otimes y = (a_1 \otimes k) \otimes n$. Также y должно быть решением уравнения $a_2 \otimes y = (a_2 \otimes k) \otimes n$. Далее возможны два варианта:

1) Для любых попарно выбранных a_1 и a_2 таких, что $a_1 \neq a_2$, значения y совпадают в обоих уравнениях. Тогда значение y единственно при любой степени принадлежности слова $\alpha_1 b^{-1}A$, и правило $A \to b\ (y)$ существует в нечетко-автоматной грамматике, и такая грамматика, согласно теореме, будет эквивалентна исходной.

2) Существуют такие a_1 и a_2, что $a_1 \neq a_2$, и значения y в этих уравнениях не совпадают, что приводит к двум разным значениям y. Таким образом, правил вида $A \to b\ (y)$ должно быть несколько, которые отличаются лишь весом, однако это невозможно, и такой нечетко-автоматной грамматики не существует.

Следствие доказано для ε-однозначной грамматики, для ε-неоднозначной следствие доказывается аналогично на основе уравнения $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$. ■

Пример. Рассмотрим грамматику с правилами $S \to bA\ (0.3)$, $A \to bA(0.4)$, $A \to \varepsilon\ (0.5)$ и $x \otimes y = \begin{cases} 0.2, \text{если } x = 0.1 \\ 0.3, \text{если } x = 0.2 \\ 0.4, \text{в прочих случаях} \end{cases}$.

Слово $\alpha b^{-1}A = \{S, bA, b^2A, b^3A, \dots\}$. S не рассматриваем в силу леммы, $\mu(bA) = 0.3$, $\mu(b^2A) =$

0.4, $\mu(b^n A) = 0.5$ для $n > 2$. В данной грамматике $k = 0.4$ и $n = 0.5$. Уравнение вида $a \otimes y = (a \otimes 0.4) \otimes 0.5$ при $a = 0.3$, $a = 0.4$ или $a = 0.5$ имеет одно и то же решение $y = R$ (второй случай), значения которых не зависят от a. Выбираем любое значение решения (например, $y = 100$), и добавляем правило $A \to b\ (100)$, а также добавляем $S \to b\ (0.4)$ согласно лемме, удалив правило $A \to \varepsilon\ (0.5)$.

Эквивалентная грамматика выглядит так: $S \to bA\ (0.3)$, $A \to bA(0.4)$, $S \to b\ (0.4)$, $A \to b\ (100)$.

Пример. $S \to bA\ (0)$, $A \to bA\ (1)$, $A \to \varepsilon\ (1)$, где $x \otimes y = \frac{x+y}{2}$. Решение уравнения $\frac{a+y}{2} = \frac{\frac{a+k}{2}+n}{2}$ всегда зависит от параметра a: $y = \frac{2n+k-a}{2}$. Кроме того, здесь $\mu(bA) = 0.5$ и $\mu(b^2 A) = 0.75$, и все значения решений будут различными. Следовательно, из данной грамматики нельзя удалить правило $A \to \varepsilon\ (1)$.

На основе теоремы и доказанных следствий, можно сформулировать критерий существования нечетко-автоматной грамматики для любой ε-нечеткой автоматной грамматики.

Критерий удаления правила $\boldsymbol{B \to \varepsilon\ (n)}$. Для того, чтобы для ε-нечетко-автоматной грамматики существовала эквивалентная нечетко-автоматная грамматики необходимо и достаточно выполнения одного из следующих условий: 1) грамматика ε-однозначна и уравнение $a \otimes y = (a \otimes k) \otimes n$ имеет хотя бы одно решение относительно y в магме $\langle R, \otimes \rangle$, не зависимое от значения a или грамматика ε-неоднозначна и уравнение $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$ имеет хотя бы одно решение относительно y в алгебре $\langle R, \otimes, \oplus \rangle$ не зависимое от значения a; 2) $\forall \alpha : \alpha b^{-1} A = S$, где $\alpha b^{-1} A$ – это сентенциальная форма из которой выводится слово α максимум за 2 шага с использованием ε-правила; 3) $\exists \alpha : \alpha b^{-1} A \neq S$ и уравнение $a \otimes y = (a \otimes k) \otimes n$ в ε-однозначной ε-нечетко-автоматной грамматике или уравнение $a \otimes y = (a \otimes l) \oplus \big((a \otimes k) \otimes n\big)$ в ε-неоднозначной грамматике имеют решения, каждое из которых зависит от параметра a, но при этом слово вида $\alpha b^{-1} A$ имеет единственную степень принадлежности для любого α, выводимого в грамматике или слово вида $\alpha b^{-1} A$ имеет такое подмножество степеней принадлежности, что хотя бы одно из значений решений указанных уравнений равны.

Замечание. Если грамматика была ε-неоднозначной, и при этом в ней нет других (не ε-правил), приводящих к неоднозначности, то после применения алгоритма, грамматика будет однозначной.

IV. Теорема об удалении аксиомы из правых частей правил

До сих пор мы не рассматривали грамматики, в которых присутствует правило $S \to \varepsilon\ (n)$, когда S присутствует в правых частях правил. Оказывается аксиому можно удалить из правых частей правил – об этом говорит следующая теорема.

Теорема (об удалении аксиомы из правых частей правил). Для любой ε-нечетко-автоматной грамматики с правилом $S \to \varepsilon\ (n)$, в которой S присутствует в правых частях правил, существует эквивалентная ε-нечетко-автоматная грамматика с отсутствующей аксиомой S в правых частях правил и обратно.

Алгоритм. Опишем алгоритм удаления аксиомы S:

1) Если грамматика содержит правило $S \to \varepsilon\ (n)$, то найти все правила вида $A \to bS\ (m)$;

2) Добавить в нетерминальный алфавит новый нетерминал B, и добавить в грамматику правила $A \to bB\ (m)$ и $B \to \varepsilon\ (n)$, а также правила вида $B \to \beta\ (k)$ такие, что в грамматике существуют правила $S \to \beta\ (k)$ и $\beta \neq \varepsilon$;

3) Удалить правила $A \to bS\ (m)$.

Доказательство. Докажем корректность алгоритма: исходная и преобразованная грамматики должны порождать одно и то же слово γ. Не теряя общности, можно считать, что в выводе этого слова до сентенциальной формы $\alpha b^{-1} A$, участвуют правила, не содержащие в правых частях аксиому, поэтому при применении алгоритма они не меняются. В противном случае, если таковые правила имеются, применим алгоритм сначала к ним.

Пусть слово γ выводится в исходной грамматике. Рассмотрим шаги вывода слова после сентенциальной формы $\alpha b^{-1} A$. Пусть $\mu(\alpha b^{-1} A) = a$, тогда $\alpha b^{-1} A \to \alpha S$ и $\mu(\alpha S) = a \otimes m$. Далее возможно два варианта.

1) Применяется ε-правило $S \to \varepsilon\ (n)$, тогда $\alpha S \to \alpha = \gamma$ и $\mu(\gamma) = a \otimes m \otimes n$. Тогда в новой грамматике выводится то же слово $\alpha b^{-1} A \to \alpha B \to \alpha = \gamma$, причем $\mu(\gamma) = a \otimes m \otimes n$.

2) Применяется правило $B \to \beta\ (k)$, тогда $\alpha S \to \alpha\beta = \gamma$, и $\mu(\gamma) = a \otimes m \otimes k$. Тогда в новой грамматике выводится то же слово $\alpha b^{-1} A \to \alpha B \to \alpha\beta = \gamma$, причем $\mu(\gamma) = a \otimes m \otimes k$.

Пусть слово γ выводится в новой грамматике. Рассмотрим шаги вывода этого слова после сентенциальной формы $\alpha b^{-1} A$.

Если $\gamma = \alpha$, то это слово в новой грамматике выводится так: $\alpha b^{-1} A \to \alpha B \to \alpha$, причем $\mu(\gamma) = a \otimes m \otimes n$. Применяя алгоритм в обратном порядке, получаем исходную грамматику с правилами $A \to bS\ (m)$ и $S \to \varepsilon\ (n)$, поэтому γ выводится в грамматике: $\alpha b^{-1} A \to \alpha S \to \alpha = \gamma$ и $\mu(\gamma) = a \otimes m \otimes n$.

Если $\gamma = \alpha\beta$, то это слово в новой грамматике выводится так: $\alpha b^{-1} A \to \alpha B \to \alpha\beta$, причем $\mu(\gamma) = a \otimes m \otimes k$. Применяя алгоритм в обратном порядке, получаем исходную грамматику с правилами $A \to bS\ (m)$ и $S \to \beta\ (k)$, поэтому γ выводится в грамматике: $\alpha b^{-1} A \to \alpha S \to \alpha\beta = \gamma$ и $\mu(\gamma) = a \otimes m \otimes k$.

Заметим также, что из слова $\alpha\beta$ может быть дальнейший вывод терминального слова нечеткого языка. Если в процессе вывода такого слова также применялось правило с аксиомой в правой части, то применим указанный алгоритм и к ним, а по только что доказанному свойству, изменение правил согласно алгоритму не меняет выводимое слово и его принадлежность – и продолжаем применять алгоритм, пока не будет выведено терминальное слово. ■

Следствие. Множество ε-нечетко-автоматных грамматик с аксиомой в правых частях правил эквивалентно множеству ε-нечетко-автоматных грамматик без аксиомы в правых частях.

Таким образом, если имеется ε-нечетко-автоматная грамматика, то сначала нужно удалить аксиомы из правых частей правил, а в дальнейшем, если критерий существования нечетко-автоматной грамматики выполняется, следует преобразовать ε-нечетко-автоматную грамматику в нечетко-автоматную. В дальнейшем будем иметь в виду только нечетко-автоматную грамматику.

V. Нечеткий детерминированный конечный автомат с фантомным состоянием

Введем в рассмотрение нечеткий детерминированный конечный автомат (НДКА) – его определение дано в статье [7]. Нечетко-автоматная грамматика тесно связана с НДКА в виде следующей теоремы.

Теорема (об эквивалентности). 1) Для каждого НДКА, где $q_0 \notin F$, существует нечетко-автоматная грамматика, порождающая тот же язык. 2) Для каждой нечетко-автоматной грамматики, не содержащей правила $S \to \varepsilon$, а также совместно правил вида $A \to aB$ и $A \to aC$ или совместно правил $A \to b\ (n)$ и $A \to bB\ (m)$ при $n \neq m$ существует НДКА, распознающий тот же язык.

Доказательство теоремы и алгоритмы преобразований приведены в этой же статье [7]. Заметим, что указанная теорема не охватывает случай наличия правила $S \to \varepsilon\ (n)$, что существенно, поскольку это исключает случай принадлежности пустого слова языку, причем просто так сделать q_0 финальным состоянием нельзя – степень принадлежности пустого слова будет неопределенной.

Рассмотрим возможные варианты решения этой проблемы.

I. Введение фантомного состояния. Нечеткий детерминированный конечный автомат с фантомным состоянием (ф-НДКА) – это восьмиместный кортеж $\langle T, Q, F, I, q_{ph}, \delta, \varphi, \otimes \rangle$, где q_{ph} – фантомное состояние, являющееся одновременно начальным и конечным, не имеющее переход ни в какое другое состояние и ни из какого-либо другого (в том числе в само себя), а также имеющее вес, в отличие от всех остальных состояний, а множество $I = \{q_0, q_{ph}\}$ – множество начальных состояний. То есть $q_{ph} \in I$ и $q_{ph} \in F$, а также $\nexists \delta(q_{ph}, q)$ и $\nexists \delta(q, q_{ph})$ для $\forall q \in Q$, а также дополнительно $\exists!\, \varphi(q_{ph}, \emptyset, q_{ph})$, где $\emptyset$ означает, что весовая функция присваивает состоянию вес указанному состоянию без перехода (точнее, пустому переходу между одним и тем же состоянием). Все остальные элементы определяются так же, как и у обычного НДКА.

Тогда правилу вида $S \to \varepsilon\ (n)$ будет соответствовать старт из фантомного состояния с весом $\varphi(q_{ph}, \emptyset, q_{ph}) = n$ и завершению в нем же без каких-либо переходов. Всем остальным правилам будет соответствовать старт из состояния q_0, как и ранее. Можно считать, что ф-НДКА действительно будет детерминированным, поскольку для каждого слова имеется единственная последовательность состояний.

Теорема (об эквивалентности нечетко-автоматных грамматик и ф-НДКА). 1) Для каждого ф-НДКА, где $q_0 \notin F$, существует нечетко-автоматная грамматика, порождающая тот же язык. 2) Для каждой нечетко-автоматной грамматики, не содержащей совместно правил вида $A \to aB$ и $A \to aC$ или совместно правил $A \to b\ (n)$ и $A \to bB\ (m)$ при $n \neq m$ существует ф-НДКА, распознающий тот же язык.

Доказательство. 1) Фантомное состояние изолировано от всего автомата, и не влияет на распознаваемый язык, кроме того, что добавляет в язык пустое слово ε со степенью принадлежности $\mu(\varepsilon) = \varphi(q_{ph}, \emptyset, q_{ph})$, и в построенную грамматику по алгоритму из теоремы об эквивалентности добавляется лишь правило $S \to \varepsilon\left(\mu\left(\varphi(q_{ph}, \emptyset, q_{ph})\right)\right)$. Если $\varphi(q_{ph}, \emptyset, q_{ph}) = 0$, то ε не принадлежит языку, и в грамматике нет правила вида $S \to \varepsilon$.

2) Для грамматики без правила $S \to \varepsilon\ (n)$ утверждение записано в виде теоремы выше. Для правила $S \to \varepsilon\ (n)$ добавляем фантомное состояние, получая ф-НДКА, где $\varphi(q_{ph}, \emptyset, q_{ph}) = n$. Изолированное состояние не оказывает никакого влияния ни на какую-либо сентенциальную форму, кроме включения ε в этот язык, если $n \neq 0$.■

Таким образом, введение фантомного состояния в автомат можно считать универсальным способом включить пустое слово в автоматный язык.

Пример. Ф-НДКА для грамматики $S \to \varepsilon\ (0.25)$, $S \to a|aA\ (0.5)$, $A \to aA|a\ (1)$ с бинарной операцией $x \otimes y = \frac{x+y}{2}$ изображен на рис. 1.

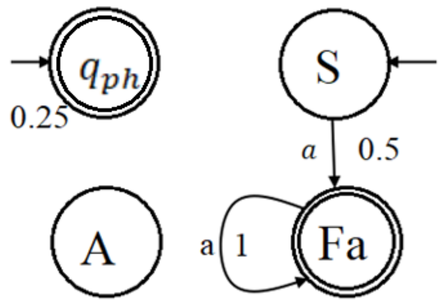

Рис. 1. Ф-НДКА для грамматики из примера

VI. Нечеткий детерминированный конечный автомат с предопределением

Второй способ решения проблемы автоматного представления заключается не введении дополнительного состояния в автомат, а в предопределении степени принадлежности стартового состояния q_0.

II. Предопределение степени принадлежности.

Автомат с предопределением определяется так же, как ранее – это семиместный кортеж $\langle T, Q, F, q_0, \delta, \varphi, \otimes \rangle$, только весовая функция дополняется следующим случаем: $\exists!\, \varphi(q_0, \emptyset, q_0) = n$, что означает присвоение начальному состоянию q_0 веса без совершения какого-либо перехода. Также следует принять соглашение, что вес q_0 присваивается лишь однократно – в тот момент, когда он не определен. Теперь, если $q_0 \in F$, то $\varepsilon \in \tilde{L}$ со степенью принадлежности $\mu(\varepsilon) = n$.

Однако, в отличие от первого способа, этот способ не универсален: существуют языки, для которых не существует НДКА с предопределенной степенью принадлежности. Покажем, что язык, заданный грамматикой $S \to \varepsilon$ (0.25), $S \to a|aA$ (0.5), $A \to aA|a$ (1) с бинарной операцией $x \otimes y = \frac{x+y}{2}$ не задается никаким автоматом с предопределенной степенью принадлежности. Поскольку $\mu(\varepsilon) = 0.25$, то $\varphi(q_0, \emptyset, q_0) = 0.25$ и $q_0 \in F$. Рассмотрим слово a^2 – в грамматике оно выводится с $\mu(a^2) = 0.75$, а в автомате будет иметь другую степень принадлежности $\mu(a^2) = 0.6875$.

Сформулируем достаточное условие эквивалентности НДКА и грамматик.

Теорема (достаточное условие эквивалентности НДКА с предопределением и нечетко-автоматных грамматик). Если для $\mu(\varepsilon) = n$ выполняется равенство $n \otimes y = y$ хотя бы для всех таких y, что $S \to \alpha \neq \varepsilon\ (y)$, то множество НДКА с предопределением и множество нечетко-автоматных грамматик с правилом $S \to \varepsilon\ (n)$, не содержащей совместно правил вида $A \to aB$ и $A \to aC$ или совместно правил $A \to b\ (n)$ и $A \to bB\ (m)$ при $n \neq m$, эквивалентны.

Доказательство. Доказательство проведем с помощью построения ф-НДКА. Поскольку множество ф-НДКА и указанных грамматик эквивалентны, тем самым мы докажем эквивалентность множеств НДКА с предопределением с множеством нечетко-автоматных грамматик, причем, поскольку нечетко-автоматные грамматики с правилом $S \to \varepsilon\ (n)$ не содержат аксиомы в правых частей правил, то, в ф-НДКА соответствующих им переходов в начальное состояние q_0 не будет.

1) Построим по НДКА с предопределением $\langle T, Q, F, q_0, \delta, \varphi, \otimes \rangle$ ф-НДКА $\langle T, Q, F, I, q_{ph}, \delta, \varphi, \otimes \rangle$. Введение новое фантомное состояние $Q \coloneqq Q \cup q_{ph}$, тогда $I = \{q_0, q_{ph}\}$, причем $q_0 \notin F$, то есть $F \coloneqq F \backslash q_0$. Добавим в φ новый переход по пустому множеству: $\varphi(q_{ph}, \emptyset, q_{ph}) = \varphi(q_0, \emptyset, q_0) = n$, удалив $\varphi(q_0, \emptyset, q_0)$.

Теперь, если существует какой-либо переход $\delta(q_k, a) = q_0$, то введем новое финальное состояние $f_a \in F$, добавим переходы вида $\delta(q_k, a) = f_a$ и для каждого перехода вида $\delta(q_0, a) = q_1$ добавим переход $\delta(f_a, a) = q_1$ со своей весовой функцией φ. Удалим переход $\delta(q_k, a) = q_0$. Все остальное остается без изменений.

Данные автоматы распознают одни и те же слова. Пустое слово теперь распознается с той же степенью принадлежности $\varphi(q_{ph}, \emptyset, q_{ph})$, а поскольку равенство $n \otimes y = y$ верно для всех y, что $S \to \alpha \neq \varepsilon\ (y)$, то есть если было $[q_0, n, \beta] \to [q_0, n \otimes y = y, \gamma]$, то стало $[q_0, *, \beta] \to [q_0, y, \gamma]$, то ни одна сентенциальная форма не поменяла свою степень принадлежности после удаления предопределения.

Теперь покажем, что добавление нового финального состояния не повлияло на сентенциальные формы и их степени принадлежности. В изначальном автомате последовательность конфигураций была такой: $\ldots \to [q_k, t_1, ab\beta] \to [q_0, t_1 \otimes t_2, b\beta] \to [q_1, t_1 \otimes t_2 \otimes t_3, \beta] \to \ldots$ или $\ldots \to [q_k, t_1, a] \to [q_0, t_1 \otimes t_2, \varepsilon]$ в случае перехода в финальное состояние. В ф-НДКА последовательности конфигураций будут следующими соответственно: $\ldots \to [q_k, t_1, ab\beta] \to [f_a, t_1 \otimes t_2, b\beta] \to [q_1, t_1 \otimes t_2 \otimes t_3, \beta] \to \ldots$ или $\ldots \to [q_k, t_1, a] \to [f_a, t_1 \otimes t_2, \varepsilon]$, так что распознаются те же самые слова с той же степенью принадлежности, а в силу того, что переход $\delta(q_k, a) = q_0$ удален, то автомат остается детерминированным.

2) Построим по ф-НДКА $\langle T, Q, F, I, q_{ph}, \delta, \varphi, \otimes \rangle$ НДКА с предопределением $\langle T, Q, F, q_0, \delta, \varphi, \otimes \rangle$. Объединим начальное состояние и фантомное состояние в одно $q_0 \coloneqq I$, причем $\varphi(q_0, \emptyset, q_0) = \varphi(q_{ph}, \emptyset, q_{ph}) = n$ и $q_0 \in F$, если $n \neq 0$ и $q_0 \notin F$, если $n = 0$. Фантомное состояние q_{ph} исключим. Все остальное остается без изменений.

Распознаваемые слова обоими автомата совпадают. Пустое слово распознается с той же степенью принадлежности n, и поскольку $n \otimes y = y$, то объединение состояний не повлияло ни на одну степень принадлежности любой сентенциальной формы, полученной из конфигурации $[q_0, *, \alpha]$: если $[q_0, *, \beta] \to [q_0, y, \gamma]$, то в НДКА с предопределением: $[q_0, n, \beta] \to [q_0, n \otimes y = y, \gamma]$ для любых $\gamma = a\beta$. Кроме того, переходов в начальное состояние здесь изначально не было.■

Замечание. В частном случае, если $\langle R, \otimes \rangle$ является моноидом, а n – ее нейтральный элемент, то НДКА с предопределением для такой нечетко-автоматной грамматики существует всегда. Однако теорема охватывает более общий случай, когда $n \otimes y = y$ в любой магме лишь для некоторых y.

Пример. Грамматика задана правилами $S \to \varepsilon(0.3)$ $S \to aA$ (0.6), $S \to bA$ (1), $A \to cB(0.1)$, $B \to aB|bB|a|b(0.9)$ с бинарной операцией $x \otimes y = \begin{cases} y, \text{если } x \leq 0.5 \\ x \cdot y, \text{в прочих случаях} \end{cases}$. В данной грамматике $n = 0.3$, а $y = 0.6$ или $y = 1$. В любом случае $0.3 \otimes y = y$, а значит, выполняется достаточное условие эквивалентности, что означает существование НДКА с предопределением. Его диаграмма состояний изображена на рис. 2.

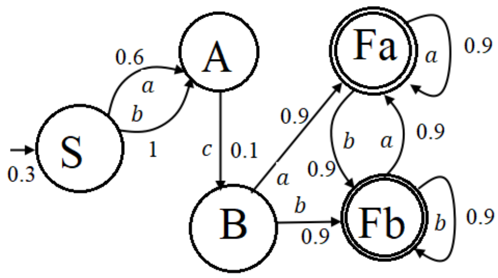

Рис. 2. НДКА с предопределением для грамматики

Утверждение. Если бинарная операция задана классическим умножением, то НДКА с предопределением для грамматики, указанной в предыдущей теореме, существует тогда и только тогда, когда существует правило $S \to \varepsilon$ (1) или $\tilde{L} = \{(\varepsilon, n)\}$ при

$n \in [0,1]$.

Доказательство. Множество весов и умножение образуют классическую группу $\langle R,\cdot\rangle$ или $\langle[0,1],\cdot\rangle$. В первом случае $n = 1$ является нейтральным элементом, а значит, выполняется достаточное условие эквивалентности НДКА с предопределением.

Рассмотрим последовательности конфигураций распознавания слова $a\alpha$ НДКА с предопределением $[q_0, n, a\alpha] \to [q_1, n\cdot y, \alpha] \to \cdots \to [q_f, n\cdot y\cdot z, \varepsilon]$, где q_f – финальное состояние. Тем временем в ф-НДКА последовательность конфигураций следующая: $[q_0,*, a\alpha] \to [q_1, y, \alpha] \to \cdots \to [q_f, y\cdot z, \varepsilon]$. Принадлежность терминальных слов должна быть одинакова: $n\cdot y\cdot z = y\cdot z$. Равенство возможно тогда и только тогда, когда или $n = 1$ (1 случай), или $y = 0$ при любом n и z, или $z = 0$ при любом n и y. Но два последних случая означают лишь то, что все терминальные слова будут распознаваться с нулевой степенью принадлежности, то есть не принадлежать языку (где-то в процессе вывода использовался $\varphi = 0$). В итоге только $\varepsilon \in_n \tilde{L}$ при $n \in [0,1]$.■

Замечание. Указанное выше условие эквивалентности НДКА с предопределением грамматикам – лишь достаточное, а значит, в случае $n \otimes y \neq y$ НДКА с предопределением может существовать, что следует из доказанного утверждения: если $n \neq 1$ и $y \neq 0$ и $z = 0$, то равенство $n \otimes y = y$ не выполняется, но для $\tilde{L} = \{(\varepsilon, n)\}$ НДКА с предопределением существует.

VII. Заключение

Таким образом, мы проанализировали несколько аспектов ε-нечетко-автоматных грамматик: во-первых, был описан алгоритм удаления аксиомы из правых частей грамматик, во-вторых, найден способ сведения ε-нечетко-автоматной грамматики к обычной нечетко-автоматной грамматики в тех случаях, когда это возможно сделать. Кроме того, отдельно была проанализирована нечетко-автоматная грамматика с правилом $S \to \varepsilon\,(n)$ и определены два способа построения автомата для такой грамматики: ф-НДКА (общий способ), и НДКА с предопределением (частный способ). Доказано, что ф-НДКА эквивалентно таким грамматикам, а для НДКА с предопределением найдено достаточное условие существования. Исходя из этого, ε-нечетко-автоматную грамматику можно представить в виде детерминированных автоматов в случаях, когда это позволяют условия доказанных теорем.

Дальнейшие исследования могут предполагать поиск необходимого условия существования НДКА с предопределением. Кроме того, вне рамок статьи остался вопрос о преобразовании ε-нечетко-автоматных грамматик, не отвечающих критерию, описанному в статье. А также в дальнейшем требуется рассмотрение случая, когда в грамматике существуют одновременно правила вида $A \to aB$ и $A \to aC$ или правила $A \to b\ (n)$ и $A \to bB\ (m)$ при $n \neq m$, ведь наличие таких правил приводит к нечеткому недетерминированному автомату в общем случае.

Благодарности

Статья написана в рамках развития направления «Математическая лингвистика» на кафедре ВМ-1 МИЭТ. Представленные доказательства теорем войдут в диссертацию. Работа продолжает серию публикаций по исследованию формальных языков (в статье [5]) и исследованию нечетких формальных языков, начатому в статье [7].

Библиография

[1] Klir G., Yuan B.: Fuzzy Sets and Fuzzy Logic. Theory and Applications. (1995), Prentice Hall

[2] Moraga, C.: An approach to fuzzy context free languages. ESTYLF08, Cuencas Minera(Mieres - Langreo), pp. 17–19 (2008)

[3] Moraga C. Some Properties of Fuzzy Languages// Computational Intelligence. Theory and Applications. International Conference 9th Fuzzy Days At: Dortmund, Germany 2006. – P.367-374

[4] Zadeh L.A.: Fuzzy Sets. Information and Control 8 (3), (1965), 338-353

[5] Козлов С. В., А. В. Светлаков А. В. О LL(1)-грамматиках, алгоритмах на них и методах их анализа в программировании // International Journal of Open Information Technologies. – 2022. – Т. 10. – № 3. – С. 30-38.

[6] Формальные языки и компиляторы : учебное пособие для вузов / А. А. Малявко. – Москва : Издательство Юрайт, 2020. – 429 с.

[7] Светлаков, А. В. О различных способах представления некоторых нечетко-автоматных языков / А. В. Светлаков, Г. А. Банару // Системы компьютерной математики и их приложения. – 2025. – № 26. – С. 270-278.

On the transformation and automata representation of ε-fuzzy automata grammars

A. V. Svetlakov

Abstract — This paper is devoted to the analysis of ε-fuzzy-automatic grammars. We prove a theorem on the removal of rules of the form B→ε (n) from such grammars, providing an algorithm, and we also prove a theorem on the removal of axioms from the right-hand sides of such grammars, similarly providing an algorithm.

In addition, the article considers grammars containing the rule S→ε (n) and a method for their automaton representation. F-NDFAs (fuzzy deterministic finite automata with a phantom state) and NDFAs with predetermination (with a predetermined degree of membership of the initial state) are introduced. The relationship between grammars and these automata is formulated as theorems, and transformation algorithms are provided.

Key words — **fuzzy-automata grammar, ε-fuzzy-automata grammar, fuzzy deterministic finite automata with phantom states, fuzzy deterministic finite automata with predetermination, fuzzy-automata language, phantom state.**

REFERENCES

[1] Klir G., Yuan B.: Fuzzy Sets and Fuzzy Logic. Theory and Applications. (1995), Prentice Hall

[2] Moraga, C.: An approach to fuzzy context free languages. ESTYLF08, Cuencas Minera(Mieres - Langreo), pp. 17–19 (2008)

[3] Moraga C. Some Properties of Fuzzy Languages// Computational Intelligence. Theory and Applications. International Conference 9th Fuzzy Days At: Dortmund, Germany 2006. – P.367-374

[4] Zadeh L.A.: Fuzzy Sets. Information and Control 8 (3), (1965), 338-353

[5] Kozlov, S. V., and A. V. Svetlakov About LL(1)-grammars, algorithms on them and methods of their analysis in programming /International Journal of Open Information Technologies, vol. 10, no. 3 (2022), pp. 30–38.

[6] Formal Languages and Compilers: A Textbook for Universities / A. A. Malyavko. – Moscow: Yurait Publishing House, 2020. – 429 pp.

[7] Svetlakov, A. V. On different ways of representing some fuzzy-automata languages / A. V. Svetlakov, G. A. Bananaru // Systems of Computer Mathematics and Their Applications. – 2025. – No. 26. – Pp. 270–278.

About the authors

A.V. Svetlakov: National Research University of Electronic Technology (MIET), Moscow (e-mail: seferlian@mail.ru).