KYBERNETIKA ČÍSLO 4, ROČNÍK 5/1969

Bounded Push Down Automata

BRANISLAV ROVAN

The bounded push down automata, a special kind of push down automata, are defined in this paper. Bounded push down automata accept exactly bounded context-free languages, defined and studied in [6], [7].

The central problem of the theory of grammars and languages is that of determining for a given class $\mathscr{E}$ of languages a class of automata which accept exactly the languages in $\mathscr{E}$. This problem was solved for regular events [1], linear languages [2], context-free languages [3], [4] and context-sensitive languages [5]. In this paper we are going to introducea utomata (the so called "bounded push down automata" – bpda) which accept bounded languages, defined and studied in [6]. By this one of the Ginsburg's open problems [7] is solved.

The basic ideas and notations of the theory of context-free languages are used just in the sense of those in [7]. From [7] is also the definition of bounded language:

Definition 1. A context-free language L (briefly "language L" in the next) on alphabet Σ is said to be *bounded*, if there are words $w_1, \ldots, w_n$ in Σ^* such that $L \subseteq \subseteq w_1^* \ldots w_n^*$.

In the next we define a special class of push down automata which will accept exactly bounded languages. bpda which accept language $L \subseteq w_1^* \ldots w_n^*$ will contain n parts which will work sequentially. The i-th part of automaton will accept for a given x in L exactly that subword of x which belongs to w_i^*.

Definition 2. *A bounded push down automaton* (bpda) is a 7-tuple $M = (K, \Sigma, \Gamma, \delta, Z_0, q_0, F \cup Q)$, where K is a finite nonempty set of states, Σ is a finite nonempty set of input symbols, Γ is a finite nonempty set of auxiliary symbols, δ is a mapping of $K \times (\Sigma \cup \{\varepsilon\}) \times \Gamma$ into finite subsets of $K \times \Gamma^*$, Z_0 in Γ is a start auxiliary symbol, q_0 in K is a start state, $F \cup Q \subseteq K$ is a set of final states, Q contains at most one element and the following properties are satisfied:

1° There exists a partition of the set $K - (\{q_0\} \cup Q) = K_1 \cup \ldots \cup K_r, (K_i \cap K_j = \emptyset$ for $i \neq j)$ such that if (t, Z_1) is in $\delta(q, a, Z)$ for q in K_i and t in K_j, then $i \leqq j$, where a is in $\Sigma \cup \{\varepsilon\}$, Z in Γ, Z_1 in Γ^*.

2° Let there be an ordering $\{q_1^{(i)}, \ldots, q_{k_i}^{(i)}\}$ of the set K_i and let the following conditions be satisfied:

A) $\delta(q_0, \varepsilon, Z_0) \subseteq \{(q_1^{(i)}, Z_0); 1 \leqq i \leqq r\}$ and $\delta(q_0, a, Z_0) = \emptyset$ for all a in Σ.

B) If $1 \leqq i \leqq r$, $1 \leqq j < k_i$, then for exactly one a in Σ there is at least one Z in Γ such that $\delta(q_j^{(i)}, a, Z) \neq \emptyset$ and is $\delta(q_j^{(i)}, a, Z) \subseteq \{(q_{j+1}^{(i)}, Z'), Z'$ in $\Gamma^*\}$.

C) If $1 \leqq i \leqq r$, then for exactly one a in Σ there is at least one Z in Γ such that $\delta(q_{k_i}^{(i)}, a, Z) \neq \emptyset$ and is $\delta(q_{k_i}^{(i)}, a, Z) \subseteq \{(q_1^{(s)}; Z'); i \leqq s \leqq r, Z'$ in $\Gamma^*\} \cup Q'$, where $Q' = \{(p, Y)\}$, p in Q, Y in Γ^* (i.e. $Q' = \emptyset$ if $Q = \emptyset$).

D) If q is in $K - (\{q_0\} \cup Q)$, Z in Γ, then $\delta(q, \varepsilon, Z) \subseteq \{(q, Z'); Z'$ in $\Gamma^*\}$.

3° $F \subseteq \{q_1^{(i)}; 1 \leqq i \leqq r\} \cup \{q_0\}$.

Definition 3. Given a bpda M let "$\vdash$" be the relation on $K \times \Sigma^* \times \Gamma^*$ defined as follows: For arbitrary q and p in K, x in $\Sigma \cup \{\varepsilon\}$, Z in Γ, w in Σ^*, α and γ in Γ^* let $(p, xw, \alpha Z) \vdash (q, w, \alpha\gamma)$ if (q, γ) is in $\delta(p, x, Z)$. Let "$\overset{*}{\vdash}$" be the reflexive and transitive closure of the relation "$\vdash$".

Definition 4. A word w is *accepted* by a bpda M, if $(q_0, w, Z_0) \overset{*}{\vdash} (d, \varepsilon, \gamma)$ for some d in $F \cup Q$ and some γ in Γ^* (i.e. there exist states $q_0, q_1, \ldots, q_n = d$ and auxiliary words $\alpha_0 = Z_0, \alpha_1, \ldots, \alpha_n = \gamma$ such that for $w = x_1 \ldots x_n$, each x_i in $\Sigma \cup \{\varepsilon\}$ holds $(q_0, x_1 \ldots x_n, \alpha_0) \vdash (q_1, x_2 \ldots x_n, \alpha_1) \vdash \ldots \vdash (q_n, \varepsilon, \alpha_n) = (d, \varepsilon, \gamma)$).

Notation. Let us denote by $T(M)$ the set of all words accepted by a bpda M.

Lemma 1. *$T(M)$ is a bounded language for each bpda M.*

Proof. It clearly follows from Def. 2 and Def. 4 that bpda are only a special kind of pda. Thus by Th. 2.5.2 of [7] $T(M)$ is a language.

Now we show that $T(M)$ is a bounded language: Consider the same notation for M as in Def. 2. Let us denote $M_i = (K, \Sigma, \Gamma, \delta_i, Z_0, q_0, F \cup Q)$, where δ_i is a restriction of the mapping δ in such sense, that $\delta_i(a, b, c) = \delta(a, b, c)$ for (a, b, c) in $(K_i \cup \{q_0\}) \times (\Sigma \cup \{\varepsilon\}) \times \Gamma$ and $\delta_i(a, b, c) = \emptyset$ otherwise. Then clearly $T(M_i) \subseteq w_i^*$ for some w_i in Σ^*. (We can obtain this w_i in this way: Let $a_1^{(i)}, \ldots, a_{k_i}^{(i)}$ be those elements of Σ for which is $\delta(q_j^{(i)}, a_j^{(i)}, Z_j) \neq \emptyset$, $1 \leqq j \leqq k_i$. Then $w_i = a_1^{(i)} \ldots a_{k_i}^{(i)}$). From the definition of the bpda it clearly follows, that $T(M) \subseteq (T(M_1) \cup \{\varepsilon\}) \cdot (T(M_2) \cup \{\varepsilon\}) \ldots (T(M_r) \cup \{\varepsilon\})$. Thus $T(M) \subseteq w_1^* \ldots w_r^*$.

Q.E.D.

In order to prove the converse, we must introduce the notion of the set $N(M)$ for given bpda M, which is similar to that one of Null (M) in [7].

Definition 5. Given a bpda M let be $N(M) = \{w \text{ in } \Sigma^*; (q_0, w, Z_0) \vdash^* (p, \varepsilon, \varepsilon), p \text{ in } F\}$, where M is as in Def. 2.

Lemma 2. *For every bounded language L there exists a bpda M such that $L = N(M)$.*

Proof. Let L be a bounded language, i.e. $L \subseteq w_1^* \dots w_n^*$, where $w_i = x_1^{(i)} \dots x_{j_i}^{(i)}$, each $x_k^{(h)}$ in Σ. Let G be a grammar generating L, i.e. $L = L(G)$, $G = (V, \Sigma, P, \sigma)$. Let us construct a bpda M in the following way:

$M = (K, \Sigma, \Gamma, \delta, \sigma, q_0, F)$, where $K = \{q_i^{(k)};\ 1 \leqq k \leqq n,\ 1 \leqq i \leqq j_k\} \cup \{q_0\}$, $\Gamma = V$, $F = \{q_1^{(i)};\ 1 \leqq i \leqq n\} \cup F_1$, $F_1 = \emptyset$ if ε is not in L and $F_1 = \{q_0\}$ otherwise. Let us define the mapping δ as follows:

$\delta(q_0, \varepsilon, \sigma) = \{(q_1^{(i)}, u_j^R);\ 1 \leqq i \leqq n,\ u_j \text{ in } V^* \text{ and } \sigma \to u_j \text{ is in } P\}$

$\delta(q_k^{(i)}, x_k^{(i)}, x_k^{(i)}) = \{(q_{k+1}^{(i)}, \varepsilon)\}$, for $1 \leqq i \leqq n$, $1 \leqq k < j_i$

$\delta(q_{j_i}^{(i)}, x_{i_i}^{(i)}, x_{j_i}^{(i)}) = \{(q_1^{(m)}, \varepsilon);\ i \leqq m \leqq n\}$, for $1 \leqq i \leqq n$

$\delta(q_r^{(s)}, \varepsilon, \xi) = \{(q_r^{(s)}, v_h^R);\ v_h \text{ in } V^*,\ \xi \to v_h \text{ is in } P\}$, for

$1 \leqq s \leqq n$, $1 \leqq r \leqq j_s$, all ξ in $V - \Sigma$.

$\delta(q, a, Z) = \emptyset$ otherwise.

It is clear that M is a bpda (with the set $Q = \emptyset$). In the next we show that $L = N(M)$.

Let x be in L, then there is a left-most derivation of x in G: $\sigma \Rightarrow u_1\xi_1 v_1 \Rightarrow u_1 u_2 \xi_2 . . v_2 \Rightarrow \dots \Rightarrow u_1 \dots u_n$, $x = u_1 \dots u_n$, each u_i in Σ^*. Then $(q_0, u_1 \dots u_n, \sigma) \vdash \vdash (q_1^{(i)}, u_1 \dots u_n, v_1^R \xi_1 u_1^R) \vdash^* (q_j^{(k)}, u_2 \dots u_n, v_1^R \xi_1) \vdash (q_j^{(k)}, u_2 \dots u_n, v_2^R \xi_2 u_2^R) \vdash^* \dots \vdash \vdash (q, \varepsilon, \varepsilon)$, where q must be in F.]Therefore, if $x = \varepsilon$ then $q = q_0$. The non-ε word x from L (i.e. from $w_i^*, \dots, w_n^*$) is expended on the input of bpda M just in the moment when M moves from some $q_{j_i}^{(i)}$ (expending the last symbol of w_i) to one of the final states $q_1^{(m)} = q$.] Thus x is in $N(M)$ and $L \subseteq N(M)$.

In order to prove the converse inclusion let x be in $N(M)$, i.e. there exist $a_0, \dots \dots, a_{s-1}$ in $\Sigma \cup \{\varepsilon\}$ and $\gamma_0, \dots, \gamma_s$ in Γ^* such that $x = a_0 \dots a_{s-1}$, $\gamma_0 = \sigma$, $\gamma_s = \varepsilon$ and $(q_0, a_0 \dots a_{s-1}, \gamma_0) \vdash (q_1^{(i)}, a_1 \dots a_{s-1}, \gamma_1) \vdash \dots \vdash (q, \varepsilon, \gamma_s) = (q, \varepsilon, \varepsilon)$, q in F. Now, let $k(0) < k(1) < \dots < k(t)$ be those nonnegative integers for which $\gamma_{k(i)} = y_i \xi_i$, ξ_i in $V - \Sigma$, y_i in V^*. (Clearly $k(0) = 0$.) From this fact it immediately follows $\gamma_{k(i)+1} = y_i z_i^R$, where z_i is in V^* and $\xi_i \to z_i$ is in P. To this sequence of moves of M corresponds the derivation $\sigma = \gamma_{k(0)}^R = \xi_0 y_0^R \Rightarrow z_0 y_0^R = a_0 \dots \dots a_{k(1)-1} \xi_1 y_1^R \Rightarrow a_0 \dots a_{k(1)-1} z_1 y_1^R = a_0 \dots a_{k(2)-1} \xi_2 y_2^R \Rightarrow \dots \Rightarrow a_0 \dots a_{s-1} = x$ in G. Thus x is in L and $L \supseteq N(M)$.

From both inclusions $L = N(M)$. Q.E.D.

Lemma 3. *For every bounded language L there exists a bpda M such that $L = T(M)$.*

Proof. By Lemma 2 there exists a bpda $M_1 = (K, \Sigma, \Gamma, \delta, Z_0, q_0, F)$ such that $L = N(M_1)$. Let us construct a bpda M as follows:

Let Z' for every Z in Γ and p be abstract symbols.
$M = (K_M, \Sigma, \Gamma_M, Z'_0, q_0, F \cup Q)$, $K_M = K \cup Q$, $K \cap Q = \emptyset$, $Q = \{p\}$,
$\Gamma_M = \Gamma \cup \{Z'; Z \text{ in } \Gamma\}$ and define δ_M in this way:

For all a in $\Sigma \cup \{\varepsilon\}$, all Z in Γ, all q in $K - Q$ let $\delta_M(q, a, Z) = \delta(q, a, Z)$
$\delta_M(q, a, Z') = \{(t, Y'\alpha); (t, Y\alpha) \text{ is in } \delta(q, a, Z), Y \text{ in } \Gamma, \alpha \text{ in } \Gamma^*\}$ if (t, ε) is not in $\delta(q, a, Z)$
$\delta_M(q, a, Z') = \{(t, Y'\alpha); (t, Y\alpha) \text{ is in } \delta(q, a, Z), Y \text{ in } \Gamma, \alpha \text{ in } \Gamma^*\} \cup \{(p, \varepsilon)\}$ if (t, ε) is in $\delta(q, a, Z)$
and let $\delta_M(q, a, Z) = \emptyset$ otherwise.
It is clear now that x is in $T(M)$ if and only if x is in $N(M_1)$. Thus $T(M) = L$.

Q.E.D.

An immediate consequence of Lemmas 1 and 3 is the following

Theorem. *A subset L of Σ^* is a bounded language if and only if there exists a bpda M such that $L = T(M)$.*

Note. The definition of bpda can be simplified in the sense of using one final state only. It is possible by a little change of the definition of δ in Def. 2 and $N(M)$. The basic idea of the proof does not change.

(Received June 4th, 1968.)

REFERENCES

[1] S. C. Kleene: Reprezentation of events in nerve sets. Automata Studies, Princeton University Press, Princeton 1956.
[2] A. L. Rosenberg: A machine realization of the linear context-free languages. Information and Control *X* (1967), 2, 175—188.
[3] N. Chomsky: Context-free grammars and push down storage. Quarterly Progress Report No. 65, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1962.
[4] R. J. Evey: The theory of applications of push down store machines. Mathematical Linguistics and Automatic Translation, Harvard Univ. Computation Lab. Report NSF-10, May, 1963.
[5] S. Y. Kuroda: Classes of languages and linear-bound automata. Information and Control *VII* (1964), 3, 360—365.
[6] S. Ginsburg, E. H. Spanier: Bounded ALGOL-like languages. Trans. Am. Math. Soc. *113* (1964), 2, 333—368.
[7] S. Ginsburg: The mathematical theory of context-free languages. McGraw-Hill, 1966.

Ohraničené zásobníkové automaty

BRANISLAV ROVAN

Jedným z hlavných problémov teórie jazykov a gramatík je: Nájsť pre danú triedu jazykov $\mathscr{E}$ triedu automatov, ktoré by príjmali práve jazyky z triedy $\mathscr{E}$. Článok sa zaoberá touto otázkou pre ohraničené bezkontextové jazyky, ktorých teóriu rozvádza S. Ginsburg v práci [7]. Uvedená je definícia ohraničeného zásobníkového automatu a veta, ktorá zaručuje, že ohraničené zásobníkové automaty príjmajú práve ohraničené jazyky. Ku každému ohraničenému jazyku v abecede Σ existujú slová $w_1, \ldots, w_n$ v abecede Σ také, že $L \subseteq w_1^* \ldots w_n^*$. Ohraničený zásobníkový automat, ktorý príjma jazyk L sa potom skladá z n častí, ktoré pracujú postupne za sebou. i-ta časť automatu bude príjmať práve tú časť slova x z L, ktorá patrí do w_i^*.

Týmto je vyriešený jeden z problémov uvedených S. Ginsburgom v [7].

Branislav Rovan, Matematický ústav SAV, Štefánikova ul. 41, Bratislava.