PROBABILISTIC TREE AUTOMATA AND CONTEXT FREE LANGUAGES

BY
M. MAGIDOR AND G. MORAN

ABSTRACT
A notion of a probabilistic tree automaton is defined and a condition is given under which it is equivalent to a usual tree automaton. A theorem about context free languages is stated.

0. Introduction.

In [1] Thatcher and Wright introduced a notion of finite automaton, whose space of work is labelled finite binary tree. This notion is an extension of the notion of finite automaton over linear tape as studied by Rabin and Scott ([2]).

In this work we define, following Rabin [3] (for the linear case), the notion of probabilistic tree automata and prove that his cut point theorem holds here as well. This actually settles a question proposed by Rabin [6].

Using the connection found by Thatcher in [4] between tree automata and production trees of context free grammars, we are able to state a cut point theorem for weighted context free languages.

1. Definitions and notations.

$P(A)$ is the power set of A and $|A|$ is the cardinality of A. A^* is the set of all finite sequences over A.

DEFINITION 1.1. The full binary tree T is the set $\{0,1\}^*$ where the tree relation is the extension relation

$$x \leqq y \text{ iff there is } z \text{ s.t. } y = xz \tag{1.1}$$

where xz means the concatenation of x and z. The two immediate successors of x are $x0$ and $x1$. The empty sequence is denoted by Λ.

---

* This research was partially supported by ONR, Contract No. N00014 69 C 0192.
Received February 4, 1970

DEFINITION 1.2. A finite tree E is a finite subset of T s.t.

$$x \in E \quad \text{and} \quad y \leqq x \Rightarrow y \in E$$

and every x in E has either two successors or none in E.

The set of maximal elements of E is the frontier of E (denoted by $Ft(E)$). Σ is a finite set.

DEFINITION 1.3. A labelled tree in the alphabet Σ is a pair (v, E) where E is a finite tree and v is a function $v: E \to \Sigma$. A labelled tree in Σ will be called a Σ-tree or just tree where no confusion is liable to arise.

The set of all Σ trees is denoted by V_Σ.

DEFINITION 1.4. Let E be a finite tree

$$E^+ = \{\Lambda\} \cup E \cdot \{0\} \cup E \cdot \{1\}$$

where the dot represents concatenation of sets ($A \cdot B$ is the set of all concatenation of an element of A with an element of B).

E^+ is clearly a finite tree.

Now we define the notion of a tree automata and it turns out that we have two different notions of automaton as to the direction of its work—from the top of the tree to its bottom or from bottom to the top. We consider trees as depicted in the following form:

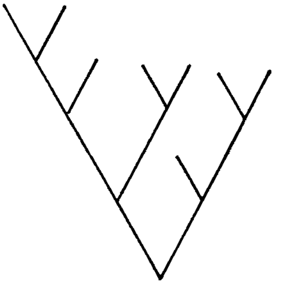

Fig. 1

DEFINITION 1.5. A sinking automaton (s.a.) over Σ is a quadruple $A = \langle S, M, S_0, F \rangle$ where S is a finite non empty set—*the set of states*

$M: S \times S \times \Sigma \to P(S)$ *is the table of moves.*

S_0 is a subset of S — *the set of initial states.*

F is a subset of S — *the set of final states.*

Let $t = (v, E) \in V_\Sigma$.

A run of A over t is a function $r\colon E^+ \to S$ s.t.

1) $r(x) \in S_0$ for $x \in Ft(E^+)$

2) $r(x) \in M(r(x0), r(x1), v(x))$ for $x \in E$.

A accepts t iff there is a run of A over t s.t. $r(\Lambda) \in F$. The set defined by A is the set of all $t \in V_\Sigma$ accepted by A and is denoted by $T(A)$.

A is a deterministic s.a. (d.s.a.) iff A is s.a. and for all $(s_1, s_2, \sigma) \in S \times S \times \Sigma$

$$|M(S_1, S_2\, \sigma)| = 1 \quad \text{and} \quad |S_0| = 1.$$

DEFINITION 1.6. A climbing automaton (c.a.) over Σ is a quadruple $\langle S, M, S_0, F\rangle$ where S, S_0, F are as in definition 1.5 and $M\colon S \times \Sigma \to P(S \times S)$ is *the automaton) table.* Let $t = (v, E) \in V_\Sigma$.

A run of A over t is a function $r\colon E^+ \to S$ s.t.

1) $r(\Lambda) \in S_0$

2) $(r(x0), r(x1)) \in M(r(x), v(x))$ for $x \in E$.

A accepts t iff there is a run r of A over t s.t. for all $x \in Ft(E^+)$ $r(x) \in F$. $T(A)$ is the set of all t's accepted by A.

The following theorem follows Thatcher and Wright.

THEOREM 1.7. The following three statements are equivalent ($B \subseteq V_\Sigma$)

1) There is a s.a. A s.t. $T(A) = B$.

2) There is a d.s.a. A s.t. $T(A) = B$.

3) There is a c.a. A s.t. $T(A) = B$.

In any of the conditions of the theorem is fulfilled we call B a regular set.

2. Algebraic characterization of regular sets.

NOTATION. Let (v, E), (u, G), $(w, H) \in V_\Sigma$ and $x \in Ft(E)$ By $[E; x; H; G]$ we mean the tree $E \cup \{x0\} \cdot H \cup \{x1\} \cdot G$. The functions v, u, w are naturally embedded in a function over $[E; x; H; G]$. The labelled tree, so obtained is called "*the adjoining of* (u, G), (w, H) *to* (v, E) *at* x" and we denote it by

$$[(v, E); x; (u, H); (w\, G)].$$

Exactly as in [2] we have an algebraic characterization of the regular subsets of V_Σ.

DEFINITION 2.1. An equivalence relation $\equiv$ over V_Σ is called *top invariant* if $t \equiv s$, and $r, z \in V_\Sigma$ $r = (v, E)$, $x \in Ft(E)$ then:

$$[r; x; t; z] \equiv [r; x; s; z] \quad \text{and} \quad [r; x; z; t] \equiv [r; x; z; s]. \tag{2.1}$$

THEOREM 2.2. *$A \subseteq V_\Sigma$ is regular iff A is the union of equivalence classes of a top invariant equivalence relation of a finite index.*

PROOF. (a) Let A be a regular, let $\boldsymbol{B}$ be the d.s.a. defining A. Define $\equiv$ by:

(2.2) $s \equiv t \Leftrightarrow$ for any $\boldsymbol{B}$ run r_1 over s and any $\boldsymbol{B}$ run r_2 over $t: r_1(\Lambda) = r_2(\Lambda)$.

$\equiv$ is easily seen to be a top invariant equivalence relation such that to any state corresponds naturally an equivalence class of $\equiv$ and vice versa, which means that there are only finitely many classes. A is the union of the classes corresponding to final states of $\boldsymbol{B}$.

(a) Let $\equiv$ be a top invariant equivalence relation of a finite index and let K be the set of its equivalence classes. Let $A = \cup N$ where $N \subseteq K$ and denote by $R(x)$ (for $x \in K$) any representative of x in V_Σ.

Define a *d*.c. by $\boldsymbol{B} = \langle K \cup \{I\}, M, I, N \rangle$ where I is an element not in K. M is defined by:

$$(2.3) \qquad M(x,y,\sigma) = [((\Lambda,\sigma)\{\Lambda\});\Lambda;R(x);R(y)]_\equiv \qquad x,y \in K$$

$$M(I,I,\sigma) = [(\sigma,\{\Lambda\})]_\equiv .$$

Define $M(x,I,\sigma)$ and $M(I,y,\sigma)$ as you like.

In (2.3) $[t]_\equiv$ is the equivalence class of t modulo $\equiv$.

A is clearly seen to be $T(\boldsymbol{B})$, because for a $\boldsymbol{B}$ run over t holds $r(\Lambda) = [t]_\equiv$.

Q.E.D.

Further extending the analogy with [2] we can find a canonical representation for the minimal d.s.a. defining a given regular set A.

This is done by regarding the top invariant relation over V_Σ:

$$(2.4) \qquad t \overset{A}{\equiv} s \Leftrightarrow (t \in A \leftrightarrow s \in A) \text{ and for any } h, g \in V_\Sigma$$

$$h = (u,H),\ x \in Ft(H) \text{ hold}$$

$$[h;x;g;t] \in A \Leftrightarrow [h;x;g;s] \in A$$

and

$$[h;x;t;y] \in A \Leftrightarrow [h;x;s;g] \in A .$$

One can prove that the d.s.a. $\boldsymbol{B}$ built as in Theorem 2.1 for $\overset{A}{\equiv}$ is up to isomorphism the unique minimal d.a.s. defining A.

3. Probabilistic automata.

Following Rabin in [2] we consider automaton whose transition table contains a probability distribution for all the possible transitions and the automaton allows an element of V_Σ iff it is probable enough that a random run will be a "good run". Again we have two kinds of automata a probabilistic climbing automaton (p.c.a.) and a probabilistic sinking automaton (p.s.a.).

DEFINITION 3.1. *A probabilistic climbing automaton over* Σ *is* an 5-tuple $\langle S, M, s_0; F, \lambda\rangle$ where S F have the same meaning as in the definition of c.a. $s_0 \in S$ is the initial state, M is the transition probability table:

$$M: S \times \Delta \times S \times S \to \{z \mid 0 \leqq z \leqq 1\}$$

(3.1)
$$\text{where} \quad \sum_{(s_1, s_2) \in S \times S} M(s, \sigma, s_1, s_2) = 1$$

λ is a real number $0 \leqq \lambda \leqq 1$ called the cut point of the automaton. Let $\boldsymbol{B}$ be the c.a $\langle S, M', s_0, F\rangle$ where $M'(s,\sigma) = S \times S$. A run of the p.c.a. over t is a run of $\boldsymbol{B}$ over t.

The probability of a run of A over $t = (v, E)$ is defined as

$$pb^A\,(r) = \prod M(s, \sigma, s_1, s_2) \tag{3.2}$$

where the multiplication runs over all 4-tuple (s, σ, s_1, s_2) such that there is $x \in E$ $v(x) = \sigma$ $r(x) = s$ and $r(x0) = s_1$ $r(x1) = s_2$ and such 4-tuples appear as many times as there are such 'x' $-$ s.

The probability of the tree $t = (v, E)$ by the p.c.a. is

$$pb^A(t) = \sum pb^A(r) \tag{3.3}$$

where the sum is over all the runs over t such that $r(Ft(E^+)) \subseteq F$. We shall omit the superscript in pb^A whenever no confusion is possible. A accepts t if $pb^A(t) > \lambda$ (λ is the cut point of A.) $T(A)$ is the subset of V_Σ accepted by A.

The example given in [3] can be easily extended here to prove that p.c.a. are stronger than c.a. (Any c.a is equivalent to some p.c.a.: choose a probability distribution which will yield the same set of runs with positive probability and take $\lambda = 0$.)

However we look for some general condition under which a p.c.a. is equivalent to c.a and as in [3] for linear probablistic automata we have

DEFINITION 3.2. The cut point of a p.c.a. A is called *isolated* if it is an isolated point of the set $\{pb^A(t) \mid t \in V_\Sigma\} \cup \{\lambda\}$. (We can assume $\lambda \neq pb^A(t)$ for all t— otherwise we "move" λ a little.)

THEOREM 3.1. *For any p.c.a. A with isolated cut point there exists a c.a. B such $T(B) = T(A)$.*

PROOF. By Theorem 2.1 it is sufficient to prove that the relation $\equiv^{T(A)}$ defined by (2.4) is of finite index for it can easily be seen that this relation fulfills all conditions imposed in Theorem 2.2 (except possibly being of finite index).

Let $t_i = (v_i, E_i)$ $1 \leqq i \leqq j$ be a set of trees inequivalent by $\equiv^{T(A)}$. We show that j is bounded. Because λ is isolated there is $\delta > 0$ such that the interval $[\lambda - \delta, \lambda + \delta]$ is exclusive of $\{pb^A(t) \mid t \in V_\Sigma\}$. Let $s_0, s_1, \cdots, s_{n-1}$ be the states of A and let A_i be the p.c.a. which is the same as A except that its initial state is s_i $(A = A_0)$. The probability vector of $t \in V_\Sigma$ is

$$pb(t) = (pb^{A_0}(t), \cdots . pb^{A_{n-1}}(t)) \tag{3.4}$$

$\vec{pb}(t) \in E^n$ — the n dimensional euclidean space. We use the norm l_1

$$\| \vec{pb}(t) \| = \sum_{i=0}^{n-1} | pb^{A_i}(t) | . \tag{3.5}$$

Let t and s be inequivalent by $\equiv^{T(A)} \cdot t = (v, E)$ and $s = (u, G)$. By (2.4) it means that either

(a) $s \in T(A)$ and $t \notin T(A)$ or conversely which implies

$$2\delta \leqq | p^{A_0}(s) - p^{A_0}(t) | \leqq \| \vec{pb}(s) - \vec{pb}(t) \|$$

or (b) there are $k = (n, K)$ $h = (l, E) \in V_\Sigma$ and $x \in Ft(E)$ such that $[h; x; r; s] \in T(A)$ and $[h; x; k; t] \notin T(A)$, or conversely or $[h; x; s; k] \in T(A)$, and $[h; x; t; k] \notin T(A)$ or conversely.

Assume the first alternative. The treatment of the other cases is similar.

For $m \in V_\Sigma$ let

$$p_i(m) = \Sigma\, pb^A(r) \text{ where the sum is over all runs over } [h; x; k; m]$$

such that

$$r(Ft(E - \{x\})^+) \subseteq F,\ r(Ft\{x0\} \cdot K^+)) \subseteq F \text{ and } r(x1) = s_i. \tag{3.4}$$

p_i is the probability of passing $x1$ with the state s_i and using only those runs which are "good" on the part of $[h; x; k; m]$ which is composed of $(E - \{x\}) \cup K$.

From this definition it is clear that p_i does not depend on m.

To go on with the proof we need the following:

LEMMA 3.2. $0 \leqq p_i \leqq 1$

$$p^A([h;x;k;m]) = \sum_{i=0}^{n-1} p_i \cdot pb^{A_i}(m).$$

PROOF. By careful observation of the notions meaning

Now $[h;x;k;s] \in T(A)$ and $[h;x;k;t] \notin T(A)$ implies

$$2\delta \leqq p^A([h;x;k;s]) - p^A([h;x;k;t])$$

(3.7)
$$= \sum_{i=0}^{n-1} p_i(pb^{A_i}(x) - pb^{A_i}(t)) \leqq \sum_{i=0}^{n-1} p_i \left| pb^{A_i}(s) - pb^{A_i}(t) \right|$$

$$\leqq \sum_{i=0}^{n-1} \left| pb^{A_i}(s) - pb^{A_i}(t) \right| = \left\| \underset{\rightarrow}{pb}(s) - \underset{\rightarrow}{pb}(t) \right\|.$$

The conclusion $2\delta \leqq \| pb(s) - pb(t) \|$ holds in all the other cases that $s \not\equiv^{T(A)} t$. Now $t_i \not\equiv {}^{T(A)} t_j$ for $i \neq j$ so $\| \underset{\rightarrow}{pb}(t_i) - \underset{\rightarrow}{pb}(t_j) \| \geqq 2\delta$ so we have j vectors in the set

(3.8)
$$\{(x_1, \cdots, x_n) \mid 0 \leqq x_i \leqq 1\} = c$$

such that the distance between any two (by the l_1 norm) is greater than 2δ. Now we use:

LEMMA 3.3. *If* $u_1, \cdots, u_k \in c$, $\|u_i - u_j\| \geqq \varepsilon, i \neq j$, *then* $k \geqq (1 + n/\varepsilon)^n$.

PROOF. We divide the cube c into $([1 + n/\varepsilon])^n$ cubes (where $[x]$ is the integral part of x) each side of each of them is less that ε/n. $\| u_i - u_j \| \geqq \varepsilon$ implies that there is at least one coordinate e such that $|u_i^e - u_j^e| \geqq \varepsilon/n$ which means that u_i and u_j cannot lie in the same cube, so $k \leqq ([1 - n/\varepsilon])^n$.

By the lemma we get that $j \leqq [1 + n/2\delta]^n$ so $T(A) = T(B)$ for some c.a.

$[1 + n/2\delta]^n$ is of course a bound for the number of states of the d.s.a. defining $T(A)$. Q.E.D.

Using the same method we can define the notions of a probabilistic sinking automata (p.s.a.) and prove the same kind of cut-point theorem for p.s.a.

We do not know whether p.s.a. is equivalent to p.c.a.

4. Weighted context free grammars

DEFINITION 4.1. A context free grammar (c.f.g.) is a quadruple.

$G=\langle\Sigma_0, \Sigma_{T_0}, \sigma, P\rangle$ where V is a finite non-empty set — the vocabulary Σ_T is a non-empty subset of Σ — the set of terminals, σ is an element of $\Sigma_N=\Sigma-\Sigma_T$ — the *initial symbol*, P is a finite non-empty subset of $\Sigma_N \times (\Sigma^* - \{\Lambda\})$, is the *set of production* rules of G.

We may write $\xi \overset{P}{\rightarrow} w$ as for$(\xi, w) \in P$ snd drop P where no confusion is possible. For $x, y \in \Sigma^*$ we write $x \overset{P}{\rightarrow} y$ iff there are $u, v, w \in \Sigma^*$, $\xi \in V_N$ s.t. $x = u\xi V$, $y = uwv$ and $\xi \overset{P}{\rightarrow} w$.

We write $x \overset{P}{\Rightarrow} y$ iff there is a sequence $x = z_0\ z_1 \cdots z_n = y$ s.t. for $0 \leqq i < n$ $z_i \overset{P}{\rightarrow} z_{i+1}$.

Such a sequence is called a derivation of y from x in . $L(G)$ the language of G is the set of all $x \in \Sigma^*$ s.t. x can be derived in G from σ.

DEFINITION 4.2. A weighted context free grammar (w.c.f.g.) is a six-tuple.

$G = \langle\Sigma, \Sigma_T, \sigma, P, w, \lambda\rangle$ where Σ, Σ_T, σ and P have the same meaning as in definition 4.1.

$w: P \rightarrow [0,1]$ is the weight function of the productions and λ is the cut point. $0 \leqq \lambda \leqq 1$. The weight of a derivation in $\langle\Sigma, \Sigma_T, \sigma, P\rangle$ $z_0, z_1, \cdots, z_n$ $(w(z_0, \cdots, z_n))$ is the product of the weights of the production used in the derivation.

$L(G)$ — the language of G is the set of all $x \in \Sigma_T^*$ s.t. x can be derived in $\langle\Sigma, \Sigma_T, \sigma, P\rangle$ by a derivation of weight greater than λ.

$W(x)$ — the weight of x is sup $W(\sigma = z_0, z_1, \cdots, z_n = x)$ where $z_0, \cdots, z_n$ is a derivation in $\langle\Sigma, \Sigma_T, \sigma, P\rangle$.

DEFINITION 4.3. λ is an isolated cut point of w.c.f.g. if λ is an isolated point of the set $\{W(x) \mid x \in \Sigma_T^*\} \cup \{\lambda\}$. Any derivation can be written in a natural way as a finite tree which is called the derivation tree (see [5]). Using the connection between the set of derivation trees of c.f.g. and regular subsets of V_Σ found by Thatcher, we can prove in the same form as in Section 3.

THEOREM 4.4. *If G is* w.c.f.g. *with isolated cutpoint, there is a* c.f.g. G' s.t. $L(G) = L(G')$.

REFERENCES

1. J. W. Thatcher and Wright, *Generalized finite automata theory*, Math. Systems Theory **2** (1968) 57–82.

2. M. O. Rabin and D. Scott, *Finite automata and their decision problems*, IBM. J. Res. Develop. (1959), 114–122.

3. M. O. Rabin, *Probabilistic automata*, Information and Control **6** (1963), 230–245.

4. J. W. Thatcher, *Characterizing Derivation Tress of Context Free Grammars Through a Generalization of Finite Automata Theory*, J. Comput. System Sci. **1** (1967), 317–322.

5. Y. Bar-Hillel, M. Perles and E. Shamir, *On formal properties of simple phrase structure grammars*, Z. Phonetik, Sprachneiss. Kommunikat. **14** (1961), 143–172.

6. M. O. Rabin, *Mathematical theory of automata*, Proc. Sympos. Appl. Math., Vol. 19, Amer. Math. Soc., Providence, R.I., 1968, pp. 153–175.

INSTITUTE OF MATHEMATICS,
THE HEBREW UNIVERSITY OF JERUSALEM