PUBLICATIONS DE L'INSTITUT MATHÉMATIQUE
Nouvelle série, tome 32 (46), 1981, pp. 137–142

ON STATE REDUCTION IN DETERMINISTIC FINITE TRANSDUCERS AND SOME PROPERTIES OF DETERMINISTIC REGULAR TRANSLATIONS

Irena Pevac

Abstract: It is shown that for every deterministic finite transducer there exists an equivalent deterministic finite transducer that has no e-moves in intermediate states. Using this result some properties of the deterministic regular translations are given.

Definition 1. A *deterministic finite transducer* or, briefly, *DFT M* s a 6-tuple $(Q, \Sigma, \Delta, \delta, q_0, F)$ where

1° Q is a finite set of *states*;

2° Σ is a finite *input alphabet*;

3° Δ is a finite *output alphabet*;

4° δ is a *mapping* from $Q \times (\Sigma \cup \{e\})$ to finite subsets of $Q \times \Delta^*$;

5° $q_0 \in Q$ is the *initial state*;

6° $F \subset Q$ is the set of *final states*;

where the mapping δ has the following properties: either for all $a \in \Sigma$, $\delta(q,a)$ has at most one element and $\delta(q, e) = \varnothing$, or $\delta(q, e)$ contains exactly one element, and for all $a \in \Sigma$, $\delta(q, a) = \varnothing$.

A configuration of M, and a binary relation $\vdash$ on configurations are defined in standard fashion as in [1].

Definition 2. The *deterministic regular translation* defined by M, denoted $\tau(M)$, is $\{(x, y) \mid (q_0, x, e) \overset{*}{\vdash} (q, e, y) \text{ for some } q \in F\}$.
Let us partition the set of states into the following three parts:
$Q_S = \{q_0\}$ the singleton whose element is the initial state;
$Q_F = F$ the set of final states;
$Q_I = Q \setminus (Q_S \cup Q_F)$ the set of intermediate states.
We define an *e-move* by the following binary relation on configurations $(q, x, y) \vdash (p, x, yy_1)$. An e-move is possible if $\delta(q, e) = (p, y_1)$.

Key words and phrases: deterministic finite transducer, deterministic regular translation, state reduction, closure properties.

We can assign a transition graph to every finite transducer. The set of nodes is determined by the set of states. A label a/y on an edge directed from a node p to a node q indicates that $\delta(p, a) \ni (q, y)$. The node of the initial states is marked with an arrow labeled by „start", and final states are encircled.

L e m m a 1. *For every DFT, there is an equivalent DFT such that each state is accessible from the initial state.*

P r o o f: Let us construct the following sequence of sets $S_1, \ldots, S_n$.

$S_1 = Q_S$,

$\vdots$

$S_{i+1} = S_i \cup \{q \mid \delta(p, a) = (q, y),\ p \in S_i,\ a \in \Sigma \cup \{e\},\ y \in \Delta^*\} \quad (i = 1, n).$

Since for all i we have $S_i \subset S_{i+1} \subset Q$, and since the set Q is finite, we get $S_n = S_{n+1}$ after at most k steps (k is the number of states of the *DFT*). The set S_n becomes the new set of states, and the states from $Q \setminus S_n$ are removed.

L e m m a 2. *For every DFT there is an equivalent DFT such that from each state it is possible to reach some final stete.*

P r o o f. Let us construct the sequence of sets $R_1, \ldots, R_m$ where

$R_1 = Q_F$

$\vdots$

$R_{i+1} = R_i \cup \{q \mid \delta(q, a) = (p, y),\ p \in R_i,\ a \in \Sigma \cup \{e\},\ y \in \Delta^*\} \ (i = 1, m).$

As in the previous case, after at most k steps we have $R_{m+1} = R_m$ and the set R_m becomes the new set of states.

Algorithms in Lemma 1. and Lemma 2. are similar to the corresponding algorithms for finite automata, and are not further discussed.

T h e o r e m 1. *For an arbitrary DFT $M = (Q, \Sigma, \Delta, \delta, q_0, F)$ there exists an equivalent DFT $M' = (Q', \Sigma, \Delta, \delta, q_0, F')$ where $Q' \subset Q$ and no intermediate state in Q' has an e-move.*

P r o o f: Without loss of generality, we shall suppose that the original *DFT* is reduced using algorithms in Lemma 1. and Lemma 2. i. e. that every state is accessible from the initial state, and from every state it is possible to reach some final state. As earlier, Q_T is the set of all intermediate states. Let Q_{Ie} be the set of all states in Q_I that have e-moves. The set Q_{Ie} is finite, since it is a subset of the finite set Q. Let $Q_{Ie} = \{p_1, p_2, \ldots, p_r\}$ and $\delta^{(0)} = \delta$. If $Q_{Ie} = \varnothing$, the work is finished. Otherwise, we apply the following algorithm:

For $i = 1, \ldots, r$ generate a new maping $\delta^{(i)}$ from mapping $\delta^{(i-1)}$ using transformation (1) defined below, for removing the state p_i.

The domain of $\delta^{(0)}$ is the Cartesian product $Q \times \Sigma \cup \{e\}$), and the domain of $\delta^{(i)}$ is $(Q \setminus \{p_1, \ldots, p_i\} \times (\Sigma \cup \{e\})$. Let us introduce the subset A_i of the Cartesian product $(Q \setminus \{p_1, \ldots, p_i\}) \times (\Sigma \cup \{e\})$ as follows:

$$A_i = \{(q, a) \mid \ \delta^{(i-1)}(q, a) = (p_i\ y) \text{ for some } y \in \Delta^*\}.$$

Card A_i is equal to the number of input edges to the state p_i, and it is positive, as the state p_i is accessible by hypothesis. The set $A_i \cup \{(p_i, e)\}$ denotes the part of the domain of the mapping $\delta^{(i-1)}$, where $\delta^{(i-1)}$ should be transformed to the mapping $\delta^{(i)}$ as follows:

$$\delta^{(i)}(q,a)=\begin{cases}(p, yz) & \text{for all } (q, a)\in A_i,\ \delta^{(i-1)}(q, a)=(p_i, y), \\ & \text{and } \delta^{(i-1)}(p_i\, e)=(p,z) \\ \varnothing & \text{for } q=p_i \\ \delta^{(i-1)}(q, a) & \text{otherwise}\end{cases} \qquad (1)$$

The transformation (1) denotes the removing of the state p_i, and concatenation of each input edge to the state p_i with the only output edge from the state p_i. As p_i is an intermediate state, it is an equivalent transformation; clearly, the transducer remains determinrstic.

Let us denote $A_i=\{(q_1, a_1), \ldots, (q_s, a_s)\}$. Graphicaly, transformation (1) is presented in Fig. 1. Some q_i could coincide mutually.

After repeating the transformation r times we get $Q'=Q\setminus Q_{Ie}$, and the mapping $\delta'=\delta^{(r)}$ in the transformed *DFT* M'. *Q.E.D.*

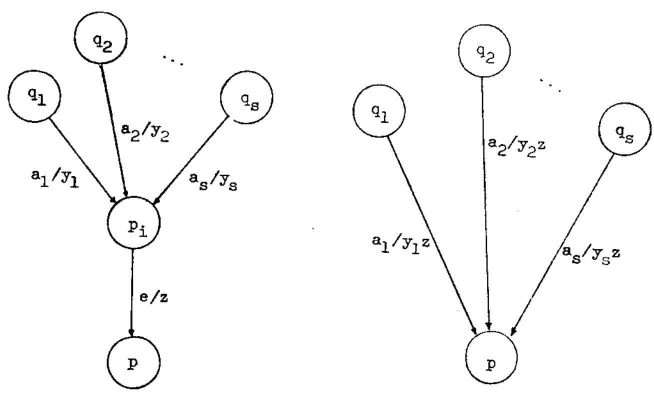

Fig. 1

The algorithm for elimination of states with e-moves, described above, cannot be used for the initial and final states. If the initial state has an *e*-move, a similar transformation given in Fig. 2. can be applied, but only for a special class of *DFT* with no edges going to state q, except for q_0.

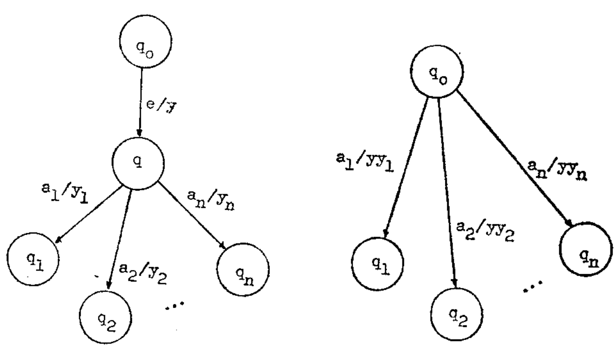

Fig. 2

The e-moves in final states produce different output words y for the same input word x, where $(x, y) \in \tau(M)$, hence they cannot be removed.

Now, we shall discuss the kind of *DRT*. We shall suppose that *DFT* is reduced by Theorem 1., and that there is no e-move in the initial state.

For an arbitrary input word x from the starting configuration (q_0, x, e) every next configuration is unique, as follows from the definition of *DFT*. After all the symbols of the input word are exhausted, and we get to some final configuration, *DFT* can still have some e-moves. Let us consider e-moves in final states.

1) The e-move from a final state to some nonfinal state $p \in Q_m$ is not allowed because it takes another e-move from p to some final state, since we must terminate in some final configuration. This contradicts the assumption that the transducer is „reduced", and has no e-moves in intermediate states.

2) Let us consider e-moves from a final to another final state. Let us suppose that we pass through n final states by e-moves. Since the out-degree is equal to one in the case of an e-move, we always get to a new state. Only from the last state in that sequence we can reach, by an e-move, some of the previous states, as Fig. 3 shows, or from

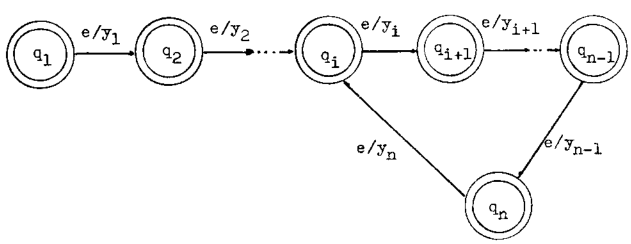

Fig. 3

q_n there is no e-move to some final state, and we have the situation given by Fig. 4.

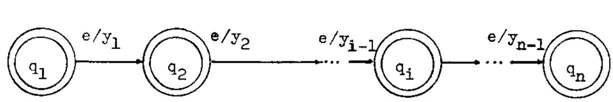

Fig. 4

If we arrive from the starting configuration to some final configuration, let us denote by (q_1, e, z) the first final configuration in which we arive from the starting configuration (q_0, x, e). As every next configuration in that sequence is unique, the output string z, emitted so far, is unique, too. Some further e-moves from a final to another final configuration are possible, after we get to the first final configuration (q_1, e, z). Again, the configurations are unique, and each time, a new string is added to the output word z.

Let us define $M(x)=\{y \mid (x, y) \in \tau(M)\}$. We have

$M(x)=\{z, zy_1\, zy_1 y_2, \ldots, zy_1 y_2 \cdots y_{i-1}, zy_1 y_2 \cdots y_{i-1}(y_i y_{i+1} \cdots y_n)^m,$

$zy_1 y_2 \cdots y_{i-1}(y_i y_{i+1} \cdots y_n)^m y_i, zy_1 y_2 \cdots y_{i-1}(y_i y_{i+1} \cdots y_n)^m y_i y_{i+1}, \ldots,$

$zy_1 y_2 \cdots y_{i-1}(y_i y_{i+1} \cdots y_n)^m y_i y_{i+1} \cdots y_{n-1} \mid m \geqslant 0, 1 \leqslant i \leqslant n\}$, if the final state q_1 is of type given in Fig. 3, and

$M(x)=\{z, zy_1, zy_1 y_2, \ldots, zy_1 y_2 \cdots y_{n-1} \mid n \geqslant 1\}$, if the final state q_1 is of type given in Fig. 4.

This leads to the following theorem:

T h e o r e m 2. *If* (x, u) *and* (x, w) *are pairs in DRT* $\tau(M)$, *then either* u *is a prefix of* w *or* w *is a prefix of* u.

P r o o f: Let us assume the contrary, i.e. let (x, u) and (x, w) both be in $\tau(M)$, and suppose that none of the strings u, w is a prefix of the other. Let U_i, $1 \leqslant i \leqslant m$, and W_i, $1 \leqslant i \leqslant n$ be sequences of successive configurations of M, such that $U_1=W_1=(q_0, x, e)$, $U_m=(p, e, u)$, and $W_n=(q, e, w)$ for some p, q in F.

Without loss of generality suppose $m \leqslant n$. If $U_m=W_m$, then u is a prefix of w, contrary to the hypothesis. Thus $U_m \neq W_m$. Finally, there is an i, $1<i \leqslant m$ such that $U_i=W_i$ and $U_j=W_j$, for all $j<i$. But then $U_{i-1} \vdash U_i$ and $U_{i-1} \vdash W_i$; this contradicts the definition of *DFT*, which implies the uniqueness of the next configuration. *Q. E. D.*

Using the previously mentioned „prefix property" we can examine the closure properties of the claas *DRT* under the operations of union, intersection, and complementation.

T h e o r e m 3. *The class of DRT is not closed under union.*

P r o o f: Let us denote by $M_i=(\{q_i, p_i\}, \{a\}\ \{0, 1\}, \delta_i, q_i, \{p_i\})$ $(i=1, 2)$ two *DFT*, where the mappings are $\delta_1(q, a)=(p_1, 01)$, and $\delta_2(q_2, a)=(p_2, 10)$, elsewhere $\delta_i=\varnothing$. The corresponding *DRT* are the following: $\tau(M)=\{(a, 01)\}$, $\tau(M_2)=\{(a, 10)\}$. $\tau(M_1) \cup \tau(M_2)=\{(a, 01), (a, 10)\}$. Pairs in $\tau(M_1) \cup \tau(M_2)$ do not have „the prefix property" from Theorem 2., and so the union is not a *DRT*.

T h e o r e m 4. *The class of DRT is not closed under complementation.*

P r o o f: Let us consider *DFT* M_1 from the proof of the previous theorem. The correspoding translation is $\tau(M_1)=\{(a, 01)\}$ while $(\tau(M_1))^c=(\{a\}^* \times \{0, 1\}^*) \setminus \{a, 01)\}$. It is easy to find two pairs in the set $(\tau(M_1))^c$ that do not have „the prefix property" from Theorem 2. So, we conclude that the complement of $\tau(M_1)$ is not a *DRT*.

T h e o r e m 5. *The class of DRT is not closed under intersection.*

P r o o f: We shall suppose that the class of *DRT* is closed under the operation of intersection. From the definition of finite transducer, which states that it is an automaton that can emit an output word y, only if the input word x is accepted by the input automaton, the result is that the regular translation is empty if and only if the input automaton language is empty. So, the emptiness problem is solvable for the class of regular translations, and consequently, for the class of *DRT*. Now, the solvability of the emptiness problem of the intersection of the two arbitrary *DRT* follows from this fact and the above assumption. But in [4] it is shown that the emptiness problem of the intersection of two *DRT* is unsolvable. Hence, the assumption is not true. So, the intersection of two *DRT* is not necesarily a *DRT*.

Acknowledgements. I would like to thank Dr Dragoš Cvetković for his careful reading of the paper, and Dr Jernej Polajnar for a helpful discussion concerning the style of proofs and presentation.

REFERENCES

[1] A. Aho, J. Ullman, *The Theory of Parsing, Translation, and Compiling*, Prentice Hall, New York, 1972.

[2] J. Hopcroft, Ullman, *Theory of fromal Languages and their Relation to Automata*, Addison-Wesley, Reading, Mass., 1969.

[3] Z. Kohavi, *Switching and finte Automata Theory*, Mc Graw-Hill, 1978.

[4] I. Marinković (Pevac), *The unsolvability of the emptiness problem of the two deterministic finte transducer mappings*, Matematički Vesnik **1(14)(29)**, (1977).

Matematički institut
Knez Mihailova 35
11000 Beograd.

(Received 24 12 1981)