Theoretical Computer Science 8 (1979) 261–263.

NOTE

A NEW PROOF OF TWO THEOREMS ABOUT RATIONAL TRANSDUCTIONS

André ARNOLD and Michel LATTEUX
UER IEEA – Informatique, Université de Lille I, 59650 Villeneuve d'Ascq, France

Communicated by Maurice Nivat
Received October 1978

Abstract. We give a new proof of Eilenberg's Cross-Section Theorem and of the theorem of decomposition of functional rational transductions.

A useful result in language theory is the "Decomposition Theorem" of Elgot and Mezei [2] (see also [1], [4] and [5]) characterizing functional rational transductions as the composition of two generalized sequential mappings (gsm). Another well-known result is the "Cross-Section Theorem" of Eilenberg [1, Theorem IX.7.1] from which Eilenberg deduces that every rational transduction contains a functional rational transduction with the same domain [1, Proposition IX.8.2]. From the Decomposition Theorem we get the following proposition:

Proposition. *For every rational transduction τ there exists two gsm σ_d and σ_g such that $\sigma_d \circ \sigma_g \subset \tau$ and* $\mathrm{dom}(\sigma_d \circ \sigma_g) = \mathrm{dom}(\tau)$.

But conversely this proposition implies the two previously mentioned theorems: if τ is functional this proposition is exactly the Decomposition Theorem and if τ is defined by $\tau(u) = h^{-1}(u) \cap \mathcal{R}$, we easily get a rational language in $\mathcal{R}$ which is in bijection with $h(\mathcal{R})$, which proves the Cross-Section Theorem.

Thus our new proof of both theorems consists in a direct proof of this proposition. The main point of this proof is a construction which is indeed dual of Schützenberger's construction of the representation of a functional rational relation by semi-monomial matrices [5]. Under its dual form this construction was recently used by Engelfriet [4] to prove a connected result about tree-transducers. The idea underlying this construction is the following.

Given a (non-deterministic) transducer defining a rational transduction τ one can extract from it a deterministic one defining a gsm σ included in τ. But then one cannot be sure that the domains of the two transducers are the same; thus in a first step one has to run through every word from right to left marking its letters with sets of states which allow to successfully go on transducing. Actually this first step looks like the first step of the algorithm for reducing finite-state automata, determining co-accessible states by determinizing reverse automata.

Proposition. *Let τ be a rational transduction from X^* into Y^* such that $\varepsilon \in \tau(\varepsilon)$. There exist a left gsm s_1 and a right one s_2 such that* $\mathrm{dom}(\tau) = \mathrm{dom}(s_1 \circ s_2)$ *and* $s_1 \circ s_2 \subseteq \tau$.

Proof. From [4], there exist an alphabet Z, a rational language $\mathscr{R} \subseteq Z^*$ and two alphabetic homomorphisms h and g such that $\hat{\tau} = \{(w, w') \mid w \in X^*, w' \in \tau(w)\} = \{(h(z), g(z)) \mid z \in \mathscr{R}\}$. Since $\varepsilon \in \tau(\varepsilon)$ we can assume that $\varepsilon \in \mathscr{R}$. Let $M = (Q, Z, *, q_0, F)$ a deterministic finite state automaton which recognizes $\mathscr{R}$. For every word u in X^* and for every subset Q' of Q let $Q'(u)$ be the set $\{q \in Q \mid \exists w \in h^{-1}(u)$ such that $q * w \in Q'\}$. Let Δ be the set $\mathscr{P}(Q) \times X \times \mathscr{P}(Q)$ and let us define the right sequential machine $M_2 = (\mathscr{P}(Q), X, \Delta, \lambda_2, \mu_2, F)$ by $\forall Q' \subseteq Q$, $\forall x \in X$, $\mu_2(Q', x) = Q'(x)$ and $\lambda_2(Q', x) = \langle Q'(x), x, Q'\rangle$. Let us also define the left generalized sequential machine $M_1 = (Q, \Delta, Y, \lambda_1, \mu_1, q_0)$ where λ_1 and μ_1 have as definition domain the set $D = \{(q, \langle Q'', x, Q'\rangle) \mid Q'' = Q'(x)$ and $q \in Q''\}$. Then λ_1 and μ_1 are defined on D by:

$$\lambda_1(q, \langle Q'', x, Q'\rangle) = g(w)$$

and

$$\mu_1(q, \langle Q'', x, Q'\rangle) = q * w,$$

where w is a word, arbitrarily chosen, such that $h(w) = x$ and $q * w \in Q'$. Let us notice that since $q \in Q'' = Q'(x)$ there exists such a word.

Let $u = x_n \cdots x_1 \in X^+$ with $x_i \in X$ for every $i \in \{1, \ldots, n\}$. It can be easily proved by induction that

(1) $\lambda_2(F, u) = \langle Q_n, x_n, Q_{n-1}\rangle \cdots \langle Q_2, x_2, Q_1\rangle\langle Q_1, x_1, F\rangle = v$

with $Q_n = F(u) = \mu_2(F, u)$.

(2) $\forall q \in Q_n, \exists w \in Z^*$ such that $q * w \in F$, $h(w) = u$ and $\lambda_1(q, v) = g(w)$.

Let us prove now that the (generalized) sequential mappings s_1, s_2 associated with the machines M_1 and M_2 satisfy the proposition.

If $u \notin \mathrm{dom}(\tau)$, from (1), $q_0 \notin F(u) = Q_n$ and by construction of M_1, $v = s_2(u) \notin \mathrm{dom}(s_1)$, which implies $u \notin \mathrm{dom}(s_1 \circ s_2)$.

If $u \in \mathrm{dom}(\tau)$, it follows from (1) that $q_0 \in F(u) = Q_n$, and, from (2), $s_1(v) = g(w)$ with $w \in \mathscr{R}$ and $h(w) = u$, thus $s_1 \circ s_2(u) = s_1(v) \in \tau(u)$.

Since $s_1 \circ s_2(\varepsilon) = \varepsilon$ we get the result.

In particular, if τ is functional, for every $u \in \mathrm{dom}(\tau)$, card $(\tau(u)) = 1$ hence $s_1 \circ s_2(x) = \tau(x)$. Therefore:

Corollary 1 (Decomposition Theorem, [1, 2, 5]). *Let τ be a functional rational transduction such that $\varepsilon \in \tau(\varepsilon)$. There exist a left gsm s_1 and a right one s_2 such that $\tau = s_1 \circ s_2$.*

Now let τ be the transduction τ from Y^* into X^*, graph of which is $\hat{\tau} = \{(h(u), u) \mid u \in \mathscr{R}\}$, where $\mathscr{R}$ is a rational language in X^* and h a homomorphism from X^* into Y^*. From the proposition there exist s_1 and s_2 such that $\mathrm{dom}(s_1 \circ s_2) = \mathrm{dom}(\tau) = h(\mathscr{R})$ and $s_1 \circ s_2 \subseteq \tau$. Thus for every $w = h(u)$ in $h(\mathscr{R})$ we get $s_1 \circ s_2(w) \in \tau(w) = h^{-1}(w) \cap \mathscr{R} \subseteq \mathscr{R}$ and $h(s_1 \circ s_2(w)) \in h(h^{-1}(h(u))) = \{w\}$. Hence $s_1 \circ s_2(h(\mathscr{R}))$ is a rational language included in $\mathscr{R}$ which is in bijection with $h(\mathscr{R})$ under h. Therefore:

Corollary 2 (Cross-Section Theorem [1]). *Let h be a homomorphism from X^* into Y^* and $\mathscr{R}$ be a rational language included in X^*. There exists a rational language $\mathscr{R}' \subseteq \mathscr{R}$ such that h is a bijection between $\mathscr{R}'$ and $h(\mathscr{R})$.*

Acknowledgement

We are grateful to Jean Berstel for pointing out that Cross-Section Theorem were a corollary of our proposition and encouraging us to write down this note.

References

[1] S. Eilenberg, *Automata, Languages and Machines*, Vol. A (Academic Press, New York, London, 1974).
[2] C.C. Elgot and J.E. Mezei, On relations defined by generalized finite automata, *IBM J. Res. Develop.* **9** (1965) 47–68.
[3] J. Engelfriet, On tree-transducers for partial functions, *Information Processing Lett.* **7** (1978) 170–172.
[4] M. Nivat, Transductions des langages de Chomsky, *Annales de l'Institut Fourier de Grenoble* **18** (1968) 339–456.
[5] M.P. Schutzenberger, Sur les relations rationnelles entre monoïdes libres. *Theor. Comput. Sci.* **3** (1976) 243–259.