Transducers

Session 4: Two Way Transducers

Aliaume LOPEZ
TA mail* Course page† Exercises page‡

March 19, 2024

1 Previous Models

Exercise 1 (Canonical Bimachines)**.** Let us recall that the production function of a rational function f can be written $\pi_f : \Sigma^* \times \Sigma \times \Sigma^* \to \Gamma^*$. Given a rational function $f : \Sigma^* \to \Gamma^*$, we can define two congruences $\simeq_l$ and $\simeq_r$ over Σ^* as follows:

$$u \simeq_l v \iff \forall x, y \in \Sigma^*, \forall a \in \Sigma, \forall w \in \Sigma^*, \pi_f(xuy, a, w) = \pi_f(xvy, a, w)$$

And similarly for $\simeq_r$.

1. Prove that $\simeq_l$ and $\simeq_r$ have finite index.
2. Construct a canonical bimachine computing f.
3. What is the complexity of the construction?
4. Can you refine the construction by first minimising the left congruence, and then the right congruence?

This construction was used in Filiot, Gauwin, and Lhote [FGL16] to prove the decidability of the following problem: given a rational function f, is it decidable whether f can be computed by a star-free bimachine?

Exercise 2 (The Great Simplification)**.** Given a rational function f, is it decidable whether there exists a Mealy Machine that computes f?

▷ Hint 1
▷ Hint 2
▷ Hint 3
▷ Hint 4
▷ Hint 5

2 Logic

Exercise 3 (Word representations)**.** Consider two ways of representing a finite word as a model: we either have the order relation $x < y$, or we have the successor relation $x = y + 1$. Show that for both ways, $\mathbb{MSO}$ gives the same expressive power. Is it true for $\mathbb{FO}$?

*ad.lopez@uw.edu.pl
†https://www.mimuw.edu.pl/~bojan/2023-2024/przeksztalcenia-automatowe-transducers
‡https://aliaumel.github.io/transducer-exercices/

Exercise 4 (Short Formulas). Prove that there exists a family of languages L_n that are defined by a formula of size $O(n)$ but such that the minimal deterministic automaton for L_n has size $\Omega(2^n)$. What about the size of an NFA$?
▷ Hint 6
▷ Hint 7
▷ Hint 8

Exercise 5 (Logic and Monoids). Let $q \in \mathbb{N}$ be a fixed quantifier rank.

1. Prove that the $\mathbb{MSO}^q$ theory of a word uw is uniquely determined by the $\mathbb{MSO}$ theory of u and w.
2. What about the $\mathbb{FO}^q$ theory?
3. Define the map $\iota \colon \Sigma^* \to \mathcal{P}(\mathbb{MSO}^q)$ by

▷ Hint 9
▷ Hint 10

3 Two Way Deterministic

Exercise 6 (Examples and non-examples). For the following functions, provide the simplest model of computation that can express them.

- The *reverse* function
- The *sort* function
- The *cycle* function, that performs a circular permutation such, for instance mapping $abcd$ to $dabc$
- The *swap* function, that swaps the first two letters of a word

▷ Hint 11

Exercise 7 (2DFTs for Languages). Prove that the class of languages recognised by deterministic two-way transducers coincides with the class of languages recognised by deterministic finite automata **using monoids**.

Exercise 8 (Forward Images?). Let f be computed by a two-way deterministic transducer with outputs, and L be a regular language. Is it true that $f(L)$ is a regular language?
▷ Hint 12

Exercise 9 (Expressiveness). Prove that 2DFT are more expressive than rational functions. What about sweeping DFTs that can only change direction at the endpoints of the input?
▷ Hint 13
▷ Hint 14

Exercise 10 (Languages and Functions). Provide a direct proof of the following inclusion of classes:

$$\mathsf{2DFA} \cdot \mathsf{Rat} \subseteq \mathsf{Rat} \cdot \mathsf{2DFA}$$

▷ Hint 15

Exercise 11 (Class inclusions). Prove that given a function f computed by a two-way deterministic transducer with outputs, it is decidable whether f is rational.

References

[FGL16] Emmanuel Filiot, Olivier Gauwin, and Nathan Lhote. "First-order definability of rational transductions: An algebraic approach". In: 2016 31st Annual ACM/IEEE Symposium on Logic in Computer Science (LICS). 2016, pp. 1–10.

A Hints

Hint 1 (Exercise 2 Use Monoids Bimachines)**.** Prove that a rational function f that satisfies $f(\varepsilon) = \varepsilon$ can be transformed into a monoid-bimachine defined by a finite monoid M, a surjective morphism $\mu \colon \Sigma^* \twoheadrightarrow M$, and a production map $\pi \colon M \times \Sigma \times M \to \Gamma^*$, whose semantics is defined as follows for all words $w \in \Sigma^*$:

$$f(w) := \prod_{uav=w} \pi(\mu(u), a, \mu(v)) \quad .$$

The production function can be generalised to subwords as follows:

$$\pi(m_l, w, m_r) := \prod_{uav=w} \pi(m_l\mu(u), a, \mu(v)m_r) \quad .$$

Using this notation $f(w) = \pi(1_M, w, 1_M)$.

Hint 2 (Exercise 2 Decompose the problem)**.** Can you decide if a letter-to-letter unambiguous NFA with outputs is computed by a Mealy Machine? Can you decide if a rational function is computed by a letter-to-letter unambiguous NFA with output?

Hint 3 (Exercise 2 What about idempotents)**.** Let $w \in \Sigma^*$ be such that $\mu(w)^2 = \mu(w)$ ($\mu(w)$ is idempotent), and $(m_l, m_r) \in M^2$. What can you say about $\pi(m_l\mu(w), w, \mu(w)m_r)$?

Hint 4 (Exercise 2 Construct Idempotents)**.** Prove using Ramsey's theorem that for every finite monoid M there exists (a computable) $N \in \mathbb{N}$ such that for all $w \in M^*$, one can compute $w = u_1u_2u_3$ such that $\mu(u_2)$ is idempotent – $\mu(u_2)^2 = \mu(u_2)$ –, $|u_1| \leq N$ and $|u_3| \leq N$.

Hint 5 (Exercise 2 Use Quantitative Pumping Arguments)**.** Assume that f is computed by a letter-to-letter unambigous NFA with outputs, then $|f(w)| = |w|$ for all $w \in \Sigma^*$. Prove that this necessary condition is also sufficient.

To that end, notice that the map $X \mapsto \pi(m_l, w^X, m_r)$ is a function from $\mathbb{N}$ to Γ^* that must be size preserving, and therefore that $|\pi(m_l\mu(w), w, \mu(w)m_r)| = |w|$. Indeed, because μ is surjective, there exist words $(x, y) \in \Sigma^*$ such that $\mu(x) = m_l$ and $\mu(y) = m_r$. Therefore, for $X \geq 3$,

$$f(xw^Xy) = \underbrace{\pi(1_M, xw, \mu(wy))}_{\alpha} \pi(\mu(xw), w, \mu(wy))^{X-2} \underbrace{\pi(\mu(xw), y, 1_M)}_{\beta} \quad .$$

Use the above equation to conclude.

Hint 6 (Exercise 4 Good languages)**.** Consider the language L_n of words of length exactly 2^n.

Hint 7 (Exercise 4 The usual trick)**.** Let $\varphi(x, y)$ be a first order formula. Prove the equivalence between the two following formulas:

1. $\psi(x, y) := \varphi(x, z) \wedge \varphi(z, y)$.
2. $\theta(x, y) := \forall s, t.(s = x \wedge t = z) \vee (s = z \wedge t = y) \Rightarrow \varphi(s, t)$.

Hint 8 (Exercise 4 Minimal Automaton)**.** How would you prove that the minimal automaton has at least 2^n states? Using the Myhill-Nerode theorem for instance?

Hint 9 (Exercise 5 Colored Logic)**.** Define a translation of usual formulas in a coloured logic, where variables are either guaranteed to be taken in u or guaranteed to be taken in w. This can be seen as an extra type system, or a sorted logic.

Prove that formulas in this typed logic are equivalent to boolean combinations of formulas that have a single type (i.e., monochromatic formulas), taking care of counting the quantifier rank of the resulting sentences.

What have you proven?

Hint 10 (Exercise 5 Aperiodicity)**.** To prove that the monoid is aperiodic in the case of $\mathbb{FO}^q$, it suffices to prove that given a first order sentence φ, and a word w, there exists $n \in \mathbb{N}$ such that $w^n \models \varphi \iff w^{n+1} \models \varphi$. We will prove the stronger statement by induction: for sentences of quantifier rank q, w^{2^q} and w^{2^q+1} have the same q-first order types.

Hint 11 (Exercise 6 Proof for the reverse using Monoids)**.** Consider a bimachine defined in terms of monoids, i.e., defined by a morphism $\mu\colon \Sigma^* \to M$, and a production function $\pi\colon M \times \Sigma \times M \to \Gamma^*$. Let e_a be the unique idempotent in the image $\{\mu(a^k) \mid k \geq 1\}$ and e_b be the unique idempotent in the image $\{\mu(b^l) \mid l \geq 1\}$.

Consider the (generalised) outputs $\alpha := \pi(e_a, a^k, e_a e_b)$ and $\beta := \pi(e_a e_b, b^l, e_b)$. It is clear that reverse$(a^{Xk} b^{Yl}) = b^{Yl} a^{Xk}$, but it is also equal to $u_0 \alpha^X u_1 \beta^Y u_2$, where $u_0, u_1, u_2 \in \Gamma^*$. By considering Y large enough, we conclude that α is b^k. Similarly, we conclude that $\beta = a^l$. However, this is absurd, since the number of a's and b's are not preserved when $X \neq Y$.

Hint 12 (Exercise 8 The answer is no)**.** What about $f(L) = \{a^n b^n \mid n \in \mathbb{N}\}$?

Hint 13 (Exercise 9 Reverse)**.** The reverse function is not rational, but can be performed using a sweeping 2DFT.

Hint 14 (Exercise 9 Reverse Map)**.** The reverse map function is not doable by a sweeping 2DFT, but can be done by a 2DFT.

Hint 15 (Exercise 10 Use a general decomposition theorem)**.** Every deterministic two-way transducer can be decomposed into a first rational function that computes the state information about the run, followed by a unfold function, that utilizes this information together with the input word to produce the input.