Answers to Questions Formulated in the Paper “On States Observability in Deterministic Finite Automata”

Tomáš Masopust
Faculty of Information Technology, Brno University of Technology
Božetěchova 2, Brno 61266, Czech Republic
masopust@fit.vutbr.cz

Abstract

This paper gives answers to questions formulated as open in the paper “On State Observability in Deterministic Finite Automata” by A. Mateescu and Gh. Păun. Specifically, it demonstrates that for all $k \geq 2$, the families of regular languages acceptable by deterministic finite automata with no more than $k$ semi-observable states, denoted by $T_k$, are anti-AFL’s, and that the family $T_1$ differs in the closure property under Kleene $+$. 

1 Introduction

In 1987, Mateescu and Păun [1] studied state observability in completely specified deterministic finite automata without non-accessible states. They classified the states of finite automata into three types: observable, semi-observable, and non-observable. A state $q$ is said to be observable if there exists a string $w$ such that $\delta(q, w)$ is a final state; otherwise, $q$ is said to be non-observable. Moreover, an observable state $q$ is said to be semi-observable if there is a state $\delta(q, a)$ that is non-observable, for some input symbol $a$. (In what follows, we assume that each deterministic finite automaton is completely specified without non-accessible states and, therefore, call them simply deterministic finite automata.) They proved that the family of all regular languages accepted by completely specified deterministic finite automata with all states being observable, denoted by $O$, forms a proper subfamily of the family of regular languages which is not closed under union, intersection, complementation, concatenation, intersection with regular sets, $\lambda$-free homomorphism, inverse homomorphism, mirror image, and right and left quotient. However, if all the regular languages are taken over a given alphabet $\Sigma$, where $\Sigma$ is minimal for them, then the family of all such regular languages, denoted by $O(\Sigma)$, is closed under union, concatenation, and left quotient (the other properties are the same as for $O$). It is also not hard to see that if the automaton has a non-observable state, then there is an equivalent automaton which has only one non-observable state (by the minimization). Thus, we have the well-known result showing that any reg-
ular language is accepted by a deterministic finite automaton with no more than one non-observable state.

On the other hand, considering the number of semi-observable states gives rise to an infinite hierarchy of families of subregular languages, denoted by $T_k$, $k \geq 0$. In addition, it is known that for $k \geq 4$, all the families $T_k$ are anti-AFL’s, i.e., they are not closed under union, concatenation, Kleene $+$, $\lambda$-free homomorphism, inverse homomorphism, and intersection with regular sets. However, the properties of (some of) the other language families were left open. Note also that in comparison with $T_k$, $k \geq 4$, $T_0$ is closed under Kleene $+$ and, therefore, is not an anti-AFL.

This paper answers these questions and proves that except for $T_1$, all the families $T_k$, $k \geq 2$, are anti-AFL’s.

2 Preliminaries and Definitions

In this paper, we assume that the reader is familiar with formal language theory (see [2]). For an alphabet (finite nonempty set) $\Sigma$, $\Sigma^*$ represents the free monoid generated by $\Sigma$, where the unit of $\Sigma^*$ is denoted by $\lambda$. Set $\Sigma^+ = \Sigma^* - \{\lambda\}$. Let $\text{REG}$ denote the family of all regular languages. For other non-specified notions and notations, the reader is referred to [2].

Let $A = (Q, \Sigma, \delta, q_0, F)$ be a completely specified deterministic finite automaton, i.e., $Q$ is a finite set of states, $\Sigma$ is an input alphabet, $\delta : Q \times \Sigma \rightarrow Q$ is a total (completely specified) transition function, $q_0 \in Q$ is the initial state, and $F \subseteq Q$ is a (possibly empty) set of final states. Let $L(A)$ denote the language accepted by $A$, i.e., $L(A) = \{w \in \Sigma^* : \delta(q_0, w) \in F\}$, where $\delta$ is, as usual, extended to be from $Q \times \Sigma^*$ to $Q$.

In addition, in this paper we assume that all the states of $Q$ are accessible, which means that for each $q \in Q$, there exists $w \in \Sigma^*$ such that $\delta(q_0, w) = q$. In what follows, the notion of a deterministic finite automaton (DFA) stands for an automaton which is completely specified, deterministic, and without non-accessible states.

A state $q \in Q$ is said to be observable if there exists a string $w \in \Sigma^*$ such that $\delta(q, w) \in F$. Otherwise, $q$ is non-observable. An observable state $q \in Q$ is said to be semi-observable if for some $a \in \Sigma$, the state $\delta(q, a)$ is non-observable. A DFA $A$ is said to be observable if all its states are observable, and the language $L(A)$ is said to be observable, too. Let $O$ denote the family of all observable regular languages. For a given DFA $A$, denote by $\text{so}(A)$ the number of semi-observable states in $A$. For each $k \geq 0$, define the language family $R_k = \{L \in \text{REG} : \text{so}(A) \leq k\}$.

An alphabet $\Sigma$ is minimal for a given language $L$ if $L \subseteq \Sigma^*$ and for each $\Sigma' \subseteq \Sigma$ we have $L - \Sigma'^* \neq \emptyset$. Let $\text{O}(\Sigma)$ and $\text{REG}(\Sigma)$ denote the families of observable regular languages and regular languages, respectively, for which $\Sigma$ is the minimal alphabet. Finally, define the language families $R_k(\Sigma)$ in an analogous way.
3 An Overview of Known Results

This section presents an overview of the main results proved in [1]. Let $L \subseteq \Sigma^*$ be a regular language, and let $\text{Init}(L) = \{w \in \Sigma^*: wy \in L \text{ for some } y \in \Sigma^*\}$. Then, the following characterization of observable languages is known.

**Theorem 1.** A regular language $L \subseteq \Sigma^*$, where $\Sigma$ is minimal for $L$, is observable if and only if $\text{Init}(L) = \Sigma^*$.

It immediately follows from this theorem that any observable language is infinite. In addition, it is known that if $L \subseteq \Sigma^*$ is finite, then $\Sigma^* - L$ is observable. However, the other inclusion does not hold. The following properties are known.

**Theorem 2.** The family $\mathcal{O}$ is closed under Kleene $\mathcal{+}$.

**Theorem 3.** The family $\mathcal{O}$ is not closed under union, intersection, complementation, concatenation, intersection with regular sets, $\lambda$-free homomorphism, inverse homomorphism, mirror image, and right and left quotient.

**Theorem 4.** Let $\Sigma$ be an alphabet with at least two symbols. Then, the family $\mathcal{O}(\Sigma)$ is not closed under intersection, complementation, homomorphism, inverse homomorphism, mirror image, and right quotient. On the other hand, it is closed under union, concatenation, and left quotient.

**Theorem 5.**
- $\mathcal{T}_0 = \mathcal{O} \cup \{\emptyset\}$,
- $\mathcal{T}_k \subseteq \mathcal{T}_{k+1}$, $k \geq 0$, and
- $\bigcup_{k \geq 0} \mathcal{T}_k = \text{REG}$.

Let $L$ be a regular language. To count the smallest $k$ such that $L \in \mathcal{T}_k$, we can use the following lemma.

**Lemma 6.** Given a language $L \in \mathcal{T}_k - \mathcal{T}_{k-1}$, the value of $k$ can be obtained algorithmically, by constructing a minimal DFA for $L$.

The following properties are also known:
- $\mathcal{T}_k$ is not closed under union, for $k \geq 2$.
- $\mathcal{T}_k$ is not closed under concatenation, for $k \geq 2$.
- $\mathcal{T}_k$ is not closed under $\lambda$-free homomorphism, for $k \geq 2$.
- $\mathcal{T}_k$ is not closed under intersection with regular sets, for $k \geq 0$.
- $\mathcal{T}_k$ is not closed under inverse homomorphism, for $k \geq 2$.
- $\mathcal{T}_k$ is not closed under Kleene $\mathcal{+}$, for $k \geq 4$.

Except for Kleene $\mathcal{+}$, where it is not solved for the families $\mathcal{T}_k$, $k = 1, 2, 3$, it remains to solve these questions for $\mathcal{T}_1$. This is done in the next section.
4 Results

This section answers the questions formulated as open in [1]. It shows that except for the family \( \mathcal{T}_1 \), all the families \( \mathcal{T}_k \) are anti-AFL’s, for all \( k \geq 2 \).

Lemma 7. Families \( \mathcal{T}_1 \) and \( \mathcal{T}_1(\Sigma) \) are not closed under union, concatenation, \( \lambda \)-free homomorphism, and inverse homomorphism.

**Proof.**

**Union:** Let \( L_1, L_2 \subseteq \{a, b\}^* \), \( L_1 = (a + b)a^* \) and \( L_2 = (a + b)b^* \). Then, \( L_1, L_2 \in \mathcal{T}_1 \) because they are accepted by DFAs \( M_1 = (\{s, r, f\}, \{a, b\}, \delta_1, s, \{f\}) \) and \( M_2 = (\{s, r, f\}, \{a, b\}, \delta_2, s, \{f\}) \), respectively, where \( \delta_1(s, a) = \delta_1(s, b) = f \), \( \delta_1(f, a) = f \), \( \delta_1(f, b) = \delta_1(r, a) = \delta_1(r, b) = r \), and \( \delta_2(s, a) = \delta_2(s, b) = f \), \( \delta_2(f, b) = f \), \( \delta_2(f, a) = \delta_2(r, a) = \delta_2(r, b) = r \). However, \( L_1 \cup L_2 \not\in \mathcal{T}_1 \) because the minimal DFA accepting it is \( M = (\{1, 2, 3, 4, 5\}, \{a, b\}, \delta, 1, \{2, 3, 4\}) \), where \( \delta(1, a) = \delta(1, b) = 2 \), \( \delta(2, a) = 3 \), \( \delta(2, b) = 4 \), \( \delta(3, a) = 3 \), \( \delta(3, b) = 5 \), \( \delta(4, a) = 5 \), \( \delta(4, b) = 4 \), and \( \delta(5, a) = \delta(5, b) = 5 \). Clearly, states 3 and 4 are semi-observable, which implies that \( L_1 \cup L_2 \not\in \mathcal{T}_1 \).

**Concatenation:** Let \( L_1 \) and \( L_2 \) be two languages over \( \{a, b\} \) defined by the following two deterministic finite automata \( A_1 \) and \( A_2 \), respectively (see Fig. 1). As it is easy to complete the automaton, the non-observable states are omitted from now on. Then,

![](image1)

**Figure 1:** DFAs \( A_1 \) and \( A_2 \).

\( L_1, L_2 \in \mathcal{T}_1 \), with \( \{a, b\} \) being their minimal alphabet, because only states 4 and 3’ are semi-observable. As the minimal DFA accepting the concatenation of these two languages \( L_1 \) and \( L_2 \) is as in Fig. 2 which has two semi-observable states, namely 4

![](image2)

**Figure 2:** The minimal DFA for \( L_1 \cdot L_2 \) with two semi-observable states, 4 and 7.

and 7, the language \( L_1 \cdot L_2 \not\in \mathcal{T}_1 \).
Figure 3: DFAs for $L$ and $h(L)$, respectively.

Note that considering two languages over different alphabets, this result can be proved more easily. Specifically, let $L_1 = a^+$ and $L_2 = b^+$. Then, $L_1, L_2 \in \mathcal{T}_0$, but $L_1 \cdot L_2 \notin \mathcal{T}_1$.

\textbf{λ-free homomorphism:} Let $L \subseteq \{a, b\}^*$ be a language defined by the following deterministic finite automaton $A$ (see Fig. 3), and let $h : \{a, b\}^* \rightarrow \{a, b\}^*$ be a homomorphism defined as $h(a) = ab$, $h(b) = b$. Then, $L \in \mathcal{T}_1$, but $h(L) \notin \mathcal{T}_1$ because the minimal DFA accepting $h(L)$ has two semi-observable states (see Fig. 3).

\textbf{Inverse homomorphism:} Let $L \subseteq \{a, b\}^*$ be a language defined by the following deterministic finite automaton $A$ (see Fig. 4), and let $h : \{a, b\}^* \rightarrow \{a, b\}^*$ be a homomorphism defined as $h(a) = aba$ and $h(b) = bab$. Then, $L \in \mathcal{T}_1$, but $h^{-1}(L) \notin \mathcal{T}_1$ because the minimal DFA accepting $h^{-1}(L)$ has two semi-observable states (see Fig. 4).

Lemma 8. Let $\Sigma$ be an alphabet. The families $\mathcal{T}_2$, $\mathcal{T}_3$, $\mathcal{T}_2(\Sigma)$, and $\mathcal{T}_3(\Sigma)$ are not closed under Kleene $\dagger$.

\textbf{Proof.} Let $L \subseteq \{a, b\}^*$ be a language defined by the following deterministic finite automaton $A$ (see Fig. 5), and let $h : \{a, b\}^* \rightarrow \{a, b\}^*$ be a homomorphism defined as $h(a) = ab$ and $h(b) = b$. Then, $L \in \mathcal{T}_1$, but $L^+ \notin \mathcal{T}_1$. The proof for $\mathcal{T}_3$ and $\mathcal{T}_3(\Sigma)$ is analogous.

Lemma 9. Families $\mathcal{T}_1$ and $\mathcal{T}_1(\Sigma)$ are closed under Kleene $\dagger$.

\textbf{Proof.} Let $L \in \mathcal{T}_1(\Sigma)$, and let $M = (Q, \Sigma, \delta, q_0, F)$ be a minimal deterministic finite automaton such that $L(M) = L$. Moreover, let $q \in Q$ be the only semi-observable state of $M$. Apply the following algorithm constructing a deterministic finite automaton $N$ such that $L(N) = L^+$:

Figure 4: DFAs for $L$ and $h^{-1}(L)$, respectively.

Figure 5: DFAs for $L$ and $L^+$, respectively.
1. Add a $\lambda$-transition from each final state to the initial state.

2. Use the common algorithm for removing $\lambda$-transitions.

3. Use the subset construction to construct a deterministic automaton $N'$.

4. Minimize the automaton $N'$ and remove all non-accessible states.

Clearly, $N' = (2^Q, \Sigma, \delta', \{q_0\}, \{X \in 2^Q : X \cap F \neq \emptyset\})$. Let $N$ denote the final automaton, i.e., $N'$ where all non-accessible states (including the adjacent transitions) are removed.

Let $a \in \Sigma$ be such that $k = \delta(q, a)$ is non-observable in $M$. Then, if $\{q\}$ is a state of $N$, then also $\{k\} = \delta'(\{q\}, a)$ is a state of $N'$. In addition, because $M$ is minimal, $\delta(k, a) = k$ for all $a \in \Sigma$, and we have that $\{k\}$ is a non-observable state. Thus, $\{q\}$ is a semi-observable state of $N$.

Assume that $\emptyset \neq X \subseteq Q$ is a state of $N$ different from $\{q\}$ (clearly, $\emptyset$ is non-accessible). Let $p \in X$ such that $p \neq q$. As $p$ is observable which is not semi-observable in $M$, there exists $w_a \in \Sigma^*$, for each $a \in \Sigma$, such that $\delta(p, aw_a) \in F$. Thus, $\delta'(X, aw_a) \cap F \neq \emptyset$ and, therefore, $X$ is an observable state which is not semi-observable.

The arguments of the previous proof also explain why the families $\mathcal{T}_k$, $k \geq 2$, are not closed under Kleene $\cup$. Let $p$ and $q$ be semi-observable states of a deterministic finite automaton accepting $L$. Then, the deterministic finite automaton accepting $L^+$ constructed as in the previous proof can have more semi-observable states than the original one, namely $\{p\}$, $\{q\}$, and $\{p, q\}$.

Acknowledgements

This work was supported by the Czech Ministry of Education under the Research Plan No. MSM 0021630528.

References

[1] Mateescu, A., Păun, G.: On states observability in deterministic finite automata. International Journal of Computer Mathematics 21(1) (1987) 17–30

[2] Salomaa, A.: Formal languages. Academic Press, New York (1973)