UMC205: Automata and Computability

Naman Mishra

January 2024

Contents

Lecture 03: Tue 09 Jan '24

0.1 A good way to construct DFAs

Suppose we have to construct a DFA for a language L over an alphabet A.

- Think of a finite number of properties of strings that you might want to keep track of. For example, "number of a's seen so far is even".
- Identify an initial property that is true of the empty string, say p_0.
- Make sure there is a rule to update the properties which are being tracked for a string wa, based purely on the properties for w and the last input a.
- The properties should imply membership in L or non-membership in L.

0.2 DFAs Formally

Definition 0.1 (DFA)**.** A *deterministic finite-state automaton* $\mathcal{A}$ over an alphabet A is a tuple (Q, s, δ, F) where

- Q is a finite set of *states*,
- $s \in Q$ is the *start state*,
- $\delta : Q \times A \to Q$ is the *transition function*,
- $F \subseteq Q$ is the set of *final states*.

For example, the first example in **??** can be written as

$$
\begin{aligned}
A &= \{a, b\} \\
Q &= \{e, o\} \\
s &= e \\
F &= \{o\}
\end{aligned}
$$

and

$$
\begin{aligned}
\delta(e, a) &= o & \qquad \delta(o, a) &= e \\
\delta(e, b) &= e & \qquad \delta(o, b) &= o
\end{aligned}
$$

We further define $\hat{\delta} : Q \times A^* \to Q$ as the extension of δ to strings.

$$
\begin{aligned}
\hat{\delta}(q, \epsilon) &= q \\
\hat{\delta}(q, wa) &= \delta(\hat{\delta}(q, w), a)
\end{aligned}
$$

Definition 0.2 (Language of a DFA)**.** The *language of a DFA $\mathcal{A}$* is

$$L(\mathcal{A}) = \left\{ w \in A^* \mid \hat{\delta}(s, w) \in F \right\}$$

0.3 Regular Languages

Definition 0.3 (Regular Language)**.** A language L is *regular* if there exists a DFA $\mathcal{A}$ over A such that $L(\mathcal{A}) = L$.

For example, the exercises we have done so far. Another example is *any* finite language.

Theorem 0.4. The class of regular languages over an alphabet is countable.

Proof. We partition the set of all DFAs over A by their number of states. For each $n \in \mathbb{N}$, there are finitely many DFAs with n states. A countable union of finite sets is countable. Thus the set of all DFAs over A is countable. Since each regular language corresponds to at least one DFA, the set of all regular languages over A is countable. □

However, we have seen that there are uncountably many languages over any alphabet. This immediately yields the following.

Corollary 0.5. There are uncountably many languages that are not regular.

Theorem 0.6 (Closure under set operations)**.** The class of regular languages is closed under union, intersection and complementation.

Proof. For complementation, simply invert the set of final states. That is, given $\mathcal{A} = (Q, s, \delta, F)$, let $\mathcal{A}' = (Q, s, \delta, Q \setminus F)$. Then $L(\mathcal{A}') = A^* \setminus L(\mathcal{A})$, since

$$\begin{aligned} w \in L(\mathcal{A}') &\iff \hat{\delta}(s, w) \in Q \setminus F \\ &\iff \hat{\delta}(s, w) \notin F \\ &\iff w \notin L(\mathcal{A}) \\ &\iff w \in A^* \setminus L(\mathcal{A}) \end{aligned}$$

For intersection and union, define the *product* of two DFAs.

Definition 0.7 (Product)**.** Given two DFAs $\mathcal{A} = (Q, s, \delta, F)$ and $\mathcal{B} = (Q', s', \delta', F')$ over the same alphabet A, the *product* of $\mathcal{A}$ and $\mathcal{B}$ is

$$\mathcal{A} \times \mathcal{B} = (Q \times Q', (s, s'), \Delta, F \times F')$$

where $\Delta((q, q'), a) = (\delta(q, a), \delta'(q', a))$.

Note that in the above definition, the extension of Δ to strings $\hat{\Delta}$ is given by

$$\hat{\Delta}((q, q'), w) = (\hat{\delta}(q, w), \hat{\delta}'(q', w))$$

This is easily proved by induction on the length of w (or structural induction on w).

$$\begin{aligned} \hat{\Delta}((q, q'), \epsilon) &= (q, q') \\ &= (\hat{\delta}(q, \epsilon), \hat{\delta}'(q', \epsilon)) \end{aligned}$$

and if

$$\hat{\Delta}((q, q'), w) = (\hat{\delta}(q, w), \hat{\delta}'(q', w))$$

then

$$\begin{aligned} \hat{\Delta}((q, q'), wa) &= \Delta(\hat{\Delta}((q, q'), w), a) \\ &= \Delta((\hat{\delta}(q, w), \hat{\delta}'(q', w)), a) \\ &= (\delta(\hat{\delta}(q, w), a), \delta'(\hat{\delta}'(q', w), a)) \\ &= (\hat{\delta}(q, wa), \hat{\delta}'(q', wa)). \end{aligned}$$

Now let $\mathcal{A}$, $\mathcal{B}$ be DFAs over A. Then $L(\mathcal{A} \times \mathcal{B}) = L(\mathcal{A}) \cap L(\mathcal{B})$, since

$$\begin{aligned} w \in L(\mathcal{A} \times \mathcal{B}) &\iff \hat{\Delta}((s, s'), w) \in F \times F' \\ &\iff (\hat{\delta}(s, w), \hat{\delta}'(s', w)) \in F \times F' \\ &\iff \hat{\delta}(s, w) \in F \wedge \hat{\delta}'(s', w) \in F' \\ &\iff w \in L(\mathcal{A}) \wedge w \in L(\mathcal{B}) \end{aligned}$$

Since $X \cup Y = \overline{\overline{X} \cap \overline{Y}}$, closure under union follows from closure under complementation and intersection.

More directly, the DFA $(Q \times Q', (s, s'), \Delta, F \times Q' \cup F' \times Q)$ accepts the language $L(\mathcal{A}) \cup L(\mathcal{B})$. $\square$

Theorem 0.8 (Closure under concatenation)**.** The class of regular languages is closed under concatenation.

Proof. $\square$

Lecture 04: Thu 11 Jan '24

0.3.1 Two Necessary Conditions for Regular Languages

In a given DFA $\mathcal{A}$ with n states, any path of length greater than n must have a loop. Let u be the string of symbols on the path from the start state to the beginning of the loop, let v be the (non-empty) string of symbols on the loop, and let w be the string of symbols on the path from the end of the loop to the final state.

Then if uvw is accepted by $\mathcal{A}$, then so is $uv^k w$ for any $k \geq 0$.

Theorem 0.9 (Pumping Lemma)**.** For any regular language L, there exists a constant k, such that for any word $t \in L$ of the form xyz with $|y| \geq k$, there exist strings u, v and w such that

(i) $y = uvw$, $v \neq \epsilon$, and

(ii) $xuv^i wz \in L$ for each $i \geq 0$.

Proposition 0.10. The language $\{a^n b^n \mid n \geq 0\}$ is not regular.

Proof. Let $k \in \mathbb{N}$. Choose $t = a^k b^k = xyz$ where $x = \epsilon$, $y = a^k$, and $z = b^k$. Let $y = uvw$ for some non-empty v. Then $v = a^j$ for some $j \geq 1$. Then $xuv^2wz = a^{k+j}b^k$, which is not in the language. Therefore, the language is not regular. $\square$

Problem 0.1. Show that $\left\{a^{2^n} \mid n \geq 0\right\}$ is not a regular language.

Solution. Let $k \in \mathbb{N}$. Choose $t = a^{2^k} = xyz$ where $x = \epsilon$, $y = a^{2^k-1}$, and $z = a$. Let $y = uvw$ for some non-empty v. Then $v = a^j$ for some $1 \leq j < 2^k$. Then $xuv^2wz = a^{2^k+j}$, which is not in the language since $2^k < 2^k + j < 2^{k+1}$.

Problem 0.2. Is the language $\{w \cdot w \mid w \in \{0,1\}^*\}$ regular?

Proof. Let $k \in \mathbb{N}$. Choose $t = 0^k 1^k 0^k 1^k = xyz$ where $x = 0^k$, $y = 1^k$, and $z = 0^k 1^k$. Let $y = uvw$ for some non-empty v. Then $v = 1^j$ for some $1 \leq j \leq k$. If j is odd, we are done. Otherwise, $xuv^2wz = 0^k 1^{k+m} 1^m 0^k 1^k$, where $j = 2m$. This is not in the language since the second half starts with a 1. □

Lecture 05: Tue 16 Jan '24

Definition 0.11. Let $L \subseteq A^*$ be a language. The *Kleene closure* of L, denoted L^*, is defined as

$$L^* = \bigcup_{n \in \mathbb{N}} L^n$$

where $L^0 = \{\epsilon\}$ and $L^{n+1} = L^n \cdot L$.

In other words,

$$L^* = \left\{s \in A^* \mid \exists\, w \in L^{\mathbb{N}} \text{ and } n \in \mathbb{N} \text{ such that } s = w_0 \cdots w_n\right\}$$

Problem 0.3. If $L \subseteq \{a\}^*$, show that L^* is regular.

Problem 0.4. Show that there exists a language $L \subseteq A^*$ such that neither L nor its complement $A^* \setminus L$ contains an infinite regular subset.

Definition 0.12 (Ultimate periodicity)**.** A subset X of $\mathbb{N}$ is said to be *ultimately periodic* if there exist $n_0 \in \mathbb{N}$, $p \in \mathbb{N}^*$ such that for all $m \geq n_0$, $m \in X$ iff $m + p \in X$.

Proposition 0.13. A subset X being ultimately periodic is equivalent to either

- there exist $n_0 \in \mathbb{N}$, $p \in \mathbb{N}^*$ such that for all $m \geq n_0$, $m \in X \implies m + p \in X$, or
-

Definition 0.14. For a language $L \subseteq A^*$, define $\mathrm{lengths}(L)$ to be $\{\#w \mid w \in L\}$.

Theorem 0.15. If L is a regular language, then $\mathrm{lengths}(L)$ is ultimately periodic.