Lesson 3: Nondeterministic finite state automata

Theme: Nondeterministic finite state automata and pumping lemma.

1 Nondeterministic finite state automata

A nondeterministic finite state automaton (NFA) is a system \( A = \langle \Sigma, Q, q_0, F, \delta \rangle \), where each component is as follows.

- \( \Sigma \) is the alphabet.
- \( Q \) is a finite set of states.
- \( q_0 \in Q \) is the initial state.
- \( F \subseteq Q \) is the set of accepting states.
- \( \delta \subseteq Q \times \Sigma \times Q \) is the transition relation.

On input word \( w = a_1 \cdots a_n \), a run of \( A \) on \( w \) is the sequence:

\[
q_0 \ a_1 \ q_1 \ a_2 \ q_2 \ \cdots \ a_n \ q_n,
\]

where \((q_i, a_{i+1}, q_{i+1}) \in \delta\), for each \( i = 0, \ldots, n-1 \).

It is called accepting run, if \( q_n \in F \). We say that \( A \) accepts \( w \), if there is an accepting run of \( A \) on \( w \). The language of all words accepted by \( A \) is denoted by \( L(A) \). A language \( L \) is an NFA language, if there is an NFA \( A \) such that \( L = L(A) \), in which, we say that the language \( L \) is accepted by \( A \), or \( A \) accepts the language \( L \).

Remark 3.1 NFA languages are closed under intersection and union. More formally, it can be stated as follows.

- For every two NFA \( A_1 \) and \( A_2 \), there is an NFA \( A' \) such that \( L(A') = L(A_1) \cap L(A_2) \).
- For every two NFA \( A_1 \) and \( A_2 \), there is an NFA \( A' \) such that \( L(A') = L(A_1) \cup L(A_2) \).

Question: Why can we not conclude that NFA languages are closed under complementation directly from the definition of NFA?

\[ \square \]

Theorem 3.2 For every NFA \( A \), there is a DFA \( A' \) such that \( L(A) = L(A') \).

In view of Theorem 3.2, we can say that a language is regular if and only if it is accepted by an NFA.

Corollary 3.3 NFA languages are closed under complement. That is, for every NFA \( A \), there is a DFA \( A' \) such that \( L(A') = \Sigma^* - L(A) \).

Theorem 3.4 Regular languages are closed under concatenation and Kleene star. More formally, it can be stated as follows.

- If \( L_1 \) and \( L_2 \) are regular languages, so is \( L_1L_2 \).
- If \( L \) is a regular language, so is \( L^* \).

*As in the case of DFA, we can define a run of \( A \) on \( w \) starting from state \( q \) as above, but starts from state \( q \).
2 Pumping lemma

In the following, for a word $w$ and an integer $n \geq 0$, $w^n$ obtained by repeating $w$ for $n$ number of times, i.e., $w \cdots w$. By default, we define $w^0 = \epsilon$.

**Lemma 3.5 (Pumping lemma)** Let $A = \langle \Sigma, Q, q_0, F, \delta \rangle$ be an NFA. Suppose there is a word $x \in L(A)$ such that $|x| \geq |Q|$. Then, the word $x$ can be divided into three parts $u, v, w$, i.e., $x = uvw$, such that for every positive integer $n \geq 0$, $uv^n w \in L(A)$.

**Lemma 3.6 (A more refined pumping lemma)** Let $A = \langle \Sigma, Q, q_0, F, \delta \rangle$ be an NFA, and let $x, y, z$ be words such that $xyz \in L(A)$ and $|y| \geq |Q|$. Then, the word $y$ can be divided into three parts $u, v, w$, i.e., $y = uvw$, such that for every positive integer $n \geq 0$, $xuv^n wz \in L(A)$.

Appendix: Concatenation and Kleene star

For two words $u$ and $v$, $u \cdot v$ denotes the word obtained by concatenating $v$ at the end of $u$. ($u \cdot v$ reads: $u$ concatenates with $v$.) By default, $u \cdot \epsilon = \epsilon \cdot u = u$. We will usually omit $\cdot$ and simply write $uv$ instead of $u \cdot v$.

In the following, let $L_1, L_2$ and $L$ be languages. We define the following operators.

\[
L_1 \cdot L_2 := \{uv \mid u \in L_1 \text{ and } v \in L_2\} \quad \text{(Concatenation)}
\]

\[
L^n := \{u_1 \cdots u_n \mid \text{each } u_i \in L\}
\]

\[
L^* := \bigcup_{n \geq 0} L^n \quad \text{(Kleene star)}
\]

As before, we usually write $L_1 L_2$ to denote $L_1 \cdot L_2$, and $L_1 L_2$ reads as $L_1$ concatenates with $L_2$. 