COL352 Lecture 6

Contents

1 Recap

Recall the informal definition of a run on NFA. We formalize this today.

2 Definitions

Definition 1

Let $N = (Q, \Sigma, \Delta, Q_0, A)$ be an NFA, and $x \in \Sigma^*$ be a string. A run of N on x is a sequence $q_0, x_1, q_1, \ldots, x_m, q_m$ such that

1. Each $q_i \in Q$
2. Each $x_i \in \Sigma \cup \{\epsilon\}$, such that $x = x_1 x_2 \ldots x_m$.
3. $q_0 \in Q_0$
4. $\forall i : (q_{i-1}, x_i, q_i) \in \Delta$

Definition 2

$q_0, x_1, \ldots, q_{m-1}, x_m, q_m$ is an accepting run if $q_m \in A$, otherwise it is a rejecting run.

Definition 3

N accepts x if there exists an accepting run of N on x.

Definition 4

$\mathcal{L}(N) = \{x \in \Sigma^* \mid N \text{ accepts } x\}$

3 DFA v/s NFA

Question 1

Is it possible to create an NFA for any regular language?

Answer. Yes. Consider the following claim.

Claim 0.1

If L is recognized by a DFA, then L is recognized by some NFA.

Proof. Suppose $D = (Q, \Sigma, \delta, q_0, A)$ recognizes the language L.
We construct $N = (Q, \Sigma, \Delta, \{q_0\}, A)$, with $\Delta = \{(q, a, q') \in Q \times \Sigma \times Q \mid q' = \delta(q, a)\}$. Then N also recognizes D. ∎

Question 2

Is it possible to create a DFA for any language that is accepted by an NFA?

Answer. Yes. Consider the following theorem.

Theorem 1

Let N be any NFA. There exists a DFA D such that $\mathcal{L}(N) = \mathcal{L}(D)$.

Proof. We shall break this into two steps.

Theorem 2

Let N be any NFA. There exists an NFA N' without ϵ transitions such that $\mathcal{L}(N) = \mathcal{L}(D)$.

Intuition: $x \in \mathcal{L}(N)$. Let $q_0 x_1 q_1 x_2 q_2 \ldots x_m q_m$ be an accepting run.
$x = x[1]x[2]\ldots x[n]$, with $x[i] \in \Sigma$. $\langle x[i] \rangle$ is a subsequence of $\langle x_i \rangle$.
For any $j \notin \{i_1, i_2, \ldots, i_n\}$, $x_j = \epsilon$, and $x_{i_k} = x[k]$. (basically compression). High-level idea - compress $x_{i_{k-1}+1} \ldots x_{i_k}$ into a single step.
Let $N = (Q, \Sigma, \Delta, Q_0, A)$. Define NFA N' as follows:

$$N' = (Q, \Sigma, \Delta', Q_0, A')$$

Here,

$$\Delta' = \{(q, a, q') \in Q \times \Sigma \times Q \mid \exists q'' \in Q \text{ such that } q'' \text{ is reachable from } q \text{ by } \epsilon \text{ transitions and } (q'', a, q') \in \Delta\}$$

(note: this corresponds to the compression in the high-level idea)

$$A' = \{q \in Q \mid \exists q' \in A \text{ such that } q' \text{ is reachable from } q \text{ by } \epsilon \text{ transitions of } \Delta\}$$

(note: this corresponds to compressing the last part of an accepting run)

Claim 2.1

$\mathcal{L}(N) \subseteq \mathcal{L}(N')$

Proof. Suppose $x \in \mathcal{L}(N)$. Then there exists an accepting run $q_0, x_1, q_1, \ldots, x_m, q_m$ of N on x.
Let $i_1 < i_2 < \cdots < i_n$ be all the indices such that $x_i \in \Sigma$ (i.e., if $j \notin \{i_1, \ldots, i_n\}$, then $x_j = \epsilon$).
Let's construct an accepting run of N' on x. Consider the sequence $q_0 x[1] q_{i_1} x[2] \ldots x[n] q_{i_n}$.
Now note that:

1. $(q_{i_{j-1}}, x[j], q_{i_j}) \in \Delta'$

   This follows directly from the definition of Δ'; indeed, in the original sequence, we have $q_{i_{j-1}} \epsilon q_{i_{j-1}+1} \epsilon \ldots \epsilon q) i_j - 1 x[j] q_{i_j}$, so using $q = q_{i_{j-1}}, q'' = q_{i_j - 1}, q' = q_{i_j}$ in the definition, we are done.

2. $q_{i_n} \in A'$

   Note that all transitions after this are ϵ transitions, and hence the accepting state q_m is reachable from q_{i_n} using ϵ transitions, and hence $q_{i_n} \in A'$.

Hence we get an accepting run of N' on x, whence we are done. ∎

Claim 2.2

$\mathcal{L}(N') \subseteq \mathcal{L}(N)$

Proof. Suppose $x \in \mathcal{L}(N')$. Then there exists an acepting run $q_0, x[1]', q_2, \ldots, x[n']', q_{n'}$ of N' on x. Note that since there are no ϵ transitions, we have $x[i]' = x[i]$, and $n = n'$.

Now consider any $(q_{i-1}, x[i], q_i)$. Then since this is in D', we must have some q'' such that q'' is reachable from q_{i-1} by ϵ transitions and $(q'', x[i], q_i) \in D$.

Consider the corresponding path from q_{i-1} to q'', say $q_{i-1}\epsilon q'_{i-1,1}\epsilon \ldots \epsilon q'' = q'_{i-1,l_{i-1}}$.

Also note that the last state q_n is an accepting state in N', hence there exists a path starting from q_n to some accepting state q'' in A consisting solely of ϵ transitions, say $q_n\epsilon q'_{n,1}\epsilon \ldots \epsilon q'_{n,l_n} = q''$.

Consider the path

$$q_0\epsilon q'_{0,1}\epsilon \ldots \epsilon q_{0,l_0}x[1]q_1\epsilon q_{1,1}\epsilon \ldots \epsilon q_{1,l_1}x[2]q_2 \ldots q_n\epsilon q'_{n,1}\epsilon \ldots \epsilon q'_{n,l_n}$$

Then the last state q'_{n,l_n} is accepting as discussed above, and all transitions are in Δ (by definition of Δ'), and the string corresponding to this run is $\epsilon^{l_0}x[1]\epsilon^{l_1}x[2] \ldots \epsilon^{l_{n-1}}x[n]\epsilon^{l_n} = x$, as needed.

Hence we get an accepting run of N on x, whence we are done. ∎

Now these two claims together show that $\mathcal{L}(N') = \mathcal{L}(N)$, whence we have completed the proof of theorem 2.

Theorem 3

Let N' be any NFA without ϵ transitions. There exists a DFA D such that $\mathcal{L}(N) = \mathcal{L}(D)$.

Proof. Let $N' = (Q, \Sigma, \Delta, Q_0, A)$.

Construct DFA $D = (2^Q, \Sigma, \delta, Q_0, A)$, where

1. $\delta : 2^Q \times \Sigma \to 2^Q$is defined as

$$\delta(R, a) = \{q \in Q \mid \exists r \in R : (r, a, q) \in \Delta\} \quad R \in 2^Q, \text{ i.e., } R \subseteq Q, a \in \Sigma$$

2. Q_0 is a subset of 2^Q, and is hence one state.
3. $\mathcal{A} = \{R \in 2^Q \mid R \cap A \neq \emptyset\}$

Claim 3.1

$\mathcal{L}(N') \subseteq \mathcal{L}(D)$

Proof. Suppose $x \in \mathcal{L}(N')$. Then there exists an accepting run $e = q_0 x[1] q_1 \dots q_{n-1} x[n]$ of N' on x.

Let $Q_0 x[1] Q_1 \dots Q_{n-1} x[n] Q_n$ be the run of D on x.

We show the following inductive claim:

Claim 3.2

$q_i \in Q_i \forall i$

Proof. Base case: for $i = 0$, this is true since the set of starting states is Q_0.
Now suppose $i > 0$, and that the inductive hypothesis holds. So we have $q_{i-1} \in Q_{i-1}$. Since $q_{i-1} \in Q_{i-1}$, and $(q_{i-1}, x[i], q_i) \in \Delta$, we have $q_i \in \delta(Q_{i-1}, x[i])$, by the definition of δ. Now since $Q_i = \delta(Q_{i-1}, x[i])$ in the run, our inductive step is complete. ■

Now using the above claim, we get $q_n \in Q_n$. Since $q_n \in A$ (as e is an accepting run), $Q_n \cap A \neq \emptyset$, whence it follows that D accepts x. ■

Claim 3.3

$\mathcal{L}(N') \supseteq \mathcal{L}(D)$

Proof. Suppose $x \in \mathcal{L}(D)$. Let $Q_0 x[1] \dots Q_{n-1} x[n] Q_n$ be the run of D on x, where $Q_n \cap A \neq \emptyset$ (which follows from the fact that this run is an accepting run).

Claim 3.4

$\forall i$: $q_i \in Q_i \implies$ there exists a run of N' on $x_1, \dots, x_i$ which ends in state q_i.

Proof. Base case: $i = 0$: Empty string's run on q_0 is an accepting run.
Now assume $i > 0$. $q_i \in Q_i = \delta(Q_{i-1}, x[i]) \implies \exists q_{i-1} \in Q_{i-1}$ such that $(q_{i-1}, x[i], q_i) \in \Delta$. By the inductibe hypothesis, there exists a run $q_0 x[1] q_1 \dots q_{i-2} x[i-1] q_{i-1}$ of N' on $x[1] \dots x[i-1]$. Consider the run $q_0 x[1] \dots q_{i-2} x[i-1] q_{i-1} x[i] q_i$. It is a run of N on $x[1] \dots x[i]$, whence we are done. ■

Using the assumption that $Q_n \cap A \neq \emptyset$, we have a state $q_n \in Q_n \cap A$, and thus there is a run of N' on x ending in state q_n, and this is an accepting run on N'. ■

Hence we have shown that theorem 3 is true. ■

Now that we have shown these two claims, we are done. ■

Theorem 4

The class of regular languages is closed under concatenation

Proof. Let L_1, L_2 be regular languages, and let $D_1 = (Q_1, \Sigma, \delta_1, q_1, A_1)$ and $D_2 = (Q_2, \Sigma, \delta_2, q_2, A_2)$ be DFAs recognizing L_1, L_2 respectively. Then the following NFA $N = (Q, \Sigma, \Delta, Q_0, A)$ recognizes $L_1 \cdot L_2$. (Assume wlog that $Q_1 \cap Q_2 = \emptyset$).

1. $Q = Q_1 \cup Q_2$.
2. $Q_0 = \{q_1\}$.
3. $A = A_2$.
4. $\Delta = \{(r_1, a, r_2) \in Q_1 \times \Sigma \times Q_1 \mid \delta_1(r_1, a) = r_2\} \cup \{(q, \epsilon, q_2) \mid q \in A_1\} \cup \{(r_1, a, r_2) \in Q_2 \times \Sigma \times Q_2 \mid \delta_2(r_1, a) = r_2\} = \Delta_1 \cup \Delta_2 \cup \Delta_3$

Claim 4.1

If $x \in L_1 \cdot L_2$ then N accepts x.

Proof. $\exists k$ such that $x[1] \dots x[k] \in L_1$, $x[k+1] \dots x[n] \in L_2$ where $n = |x|$.

Let $r_i = \hat{\delta}_1(q_1, x[1] \dots x[i])$, $s_j = \hat{\delta}_2(q_2, x[k+1] \dots x[k+j])$.

Look at the run $r_0 x[1] r_1 \dots r_{k-1} x[k] r_k \epsilon s_0 x[k+1] s_1 \dots s_{n-k}$.

Note that the first k transitions are in Δ_1, the last $n-k$ transitions are in Δ_3, and the middle transition is in Δ_2, so this is indeed a run.
Now that s_{n-k} is in A_2 since s_{n-k} is the end of an accepting run of $x[k+1] \dots x[n]$ on D_2, so s_{n-k} is an accepting state of N as well. ■

Claim 4.2

If N accepts x, then $x \in L_1 \cdot L_2$.

Proof. Since N accepts x, there must be a path from q_1 to a state in A_2, corresponding to a run in N.
Suppose the run is $a_0 x_1 a_1 \dots a_{m-1} x_m$, where $x = x_1 x_2 \dots x_m$. Consider the largest i such that $a_i \in Q_1$. Clearly, since the last state is in A_2 which is disjoint with Q_1, we can't have $i = m$. Hence by the maximality of i, we have $a_{i+1} \in Q_2$.

Suppose there is a $j < i$ such that $a_j \in Q_2$. Consider the largest such j. Then we have $a_{j+1} \in Q_1$ since if $j = i - 1$ then we know that $a_i \in Q_1$ and otherwise we can invoke the maximality of j. This implies there is a transition of the form (q, a, q') where $q \in Q_2$ and $q' \in Q_1$, which is impossible. Hence, for all $j \le i$, we have $a_j \in Q_1$. By a similar argument, we have for all $j > i$, $a_j \in Q_2$.

Using this fact, we know that $x_{i+1} = \epsilon$, and all the other x_js are in Σ, by the definition of Δ.

Now consider the strings $S_1 = x_1 x_1 \dots x_i$ and $S_2 = x_{i+2} \dots x_m$. Since $x_{i+1} = \epsilon$, we have $x = x_1 x_1 \dots x_i x_{i+2} \dots x_m = S_1 \cdot S_2$.

Now we claim that S_1 is accepted by D_1 and S_2 is accepted by D_2.

1. S_1 is accepted by D_1

Proof. We prove this by induction.

Claim 4.3

$\hat{\delta}_1(a_0, x_1 \dots x_j) = a_j \quad \forall j \le i$

Proof. Base case is when $j = 0$, which is true since the empty string ends up in a_0 by the definition of $\hat{\delta}_1$. Now suppose $j > 0$. Then we have, by the inductive hypothesis, that $\hat{\delta}_1(a_0, x_1 \dots x_{j-1}) = a_{j-1}$. Now note that $\hat{\delta}_1(a_0, x_1 \dots x_j) = \delta(\hat{\delta}(a_0, x_1 \dots x_{j-1}), x_j) = \delta(a_{j-1}, x_j)$. This is clearly a_j, since $(a_{j-1}, x_j, a_j) \in \Delta_1$. Hence we have completed the inductive step and are done. ■

Using the result for $j = i$, we can see that $\hat{\delta}_1(a_0, S_1) = a_i$. Now note that since $(a_i, x_{i+1}, a_{i+1}) \in \Delta$, with $a_i \in Q_1, a_{i+1} \in Q_2$, we must have $(a_i, x_{i+1}, a_{i+1}) \in \Delta_2$, so $a_i \in A_1$, so D_1 accepts S_1.

■

2. S_2 is accepted by D_2

Proof. Note that from the last line of the previous proof, we have that $a_{i+1} = q_2$ as well. Hence a_{i+1} is the starting state of D_2.

Claim 4.4

$\hat{\delta}_2(q_2, x_{i+2} \dots x_{i+k+1}) = a_{i+k+2} \quad \forall i + 2 \le i + k + 2 \le m$

Proof. The proof is analogous to the proof of the claim in the previous point. ■

Using the result for $k = m - i - 2$, we have $\hat{\delta}_2(q_2, S_2) = a_m$. Since the run of x considered above is accepting on N, $a_m \in A = A_2$, so D_2 accepts S_2 as well. ■

Since S_1 is accepted by D_1, $S_1 \in L_1$. Similarly, $S_2 \in L_2$. Hence $x = S_1 \cdot S_2 \in L_1 \cdot L_2$.

■

Now that we have shown these claims, we are done. ■