Finite Automata and Regular Languages (part II)

Prof. Dan A. Simovici

UMB
1 Nondeterministic Automata
Definition

A nondeterministic finite automaton (ndfa) is a quintuple $\mathcal{M} = (A, Q, \delta, q_0, F)$, where $A$ is the input alphabet of $\mathcal{M}$, $Q$ is a finite set of states, $\delta : Q \times A \rightarrow \mathcal{P}(Q)$ is the transition function, $q_0 \in Q$ is the initial state, and $F \subseteq Q$ is the set of final states of $\mathcal{M}$. We assume $A \cap Q = \emptyset$. 
Example

Consider the ndfa

\[ M = (\{a, b\}, \{q_0, q_1, q_2, q_2, q_4\}, \delta, q_0, \{q_1, q_3\}) \]

whose transition function is defined by the table:

<table>
<thead>
<tr>
<th>Input</th>
<th>State</th>
<th>State</th>
<th>State</th>
<th>State</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>q_0</td>
<td>q_1</td>
<td>q_2</td>
<td>q_3</td>
<td>q_4</td>
</tr>
<tr>
<td>a</td>
<td>{q_1, q_2}</td>
<td>\emptyset</td>
<td>{q_3}</td>
<td>\emptyset</td>
<td>\emptyset</td>
</tr>
<tr>
<td>b</td>
<td>{q_0}</td>
<td>{q_3}</td>
<td>{q_4}</td>
<td>\emptyset</td>
<td>\emptyset</td>
</tr>
</tbody>
</table>

Note the presence of pairs \((q, a)\) such that \(\delta(q, a) = \emptyset\). We refer to such pairs as **blocking situations** of \(M\). For instance, \((q_1, a)\) is a blocking situation of \(M\).
Extending the transition function for an ndfa

As in the case of the dfa, we can extend the ndfa’s transition function \( \delta \), defined on single characters, to \( \delta^* \), defined on words. Starting from the transition function \( \delta \), we define the function \( \delta^* : Q \times A^* \rightarrow \mathcal{P}(Q) \) as follows:

\[
\begin{align*}
\delta^*(q, \lambda) &= \{q\} \\
\delta^*(q, xa) &= \bigcup_{q' \in \delta^*(q, x)} \delta(q', a)
\end{align*}
\]
Graphs of ndfas

- If $\mathcal{M} = (A, Q, \delta, q_0, F)$ is an ndfa, then the graph of $\mathcal{M}$ is the marked, directed multigraph $\mathcal{G}(\mathcal{M})$, whose set of vertices is the set of states $Q$.

- The set of edges of $\mathcal{G}(\mathcal{M})$ consists of all pairs $(q, q')$ such that $q' \in \delta(q, a)$ for some $a \in A$; an edge $(q, q')$ is labeled by the input symbol $a$, where $q' \in \delta(q, a)$.

- The initial state $q_0$ is denoted by an incoming arrow with no source, and the final states are circled.

If $q' \in \delta^*(q, x)$, then there is a path in the graph $\mathcal{G}(\mathcal{M})$ labeled by $x$ that leads from $q$ to $q'$. 
Comparing dfas and ndfas

- In the graph of a dfa $M = (A, Q, \delta, q_0, F)$ you must have exactly one edge emerging from every state $q$ and for every input symbol $a \in A$.

- In the graph of an ndfa $M = (A, Q, \delta, q_0, F)$ you may have states where no arrow emerges, or states where several arrow labeled with the same symbol emerge. Also, not every symbol needs to appear as a label of an emerging edge.
Definition

The language accepted by the ndfa $M = (A, Q, \delta, q_0, F)$ is

$$L(M) = \{ x \in A^* \mid \delta^*(q_0, x) \cap F \neq \emptyset \}.$$  

In other words, $x \in L(M)$ if there exists a path in the graph of $M$ labeled by $x$ that leads from the initial state into one of the final states. Note that it is not necessary that all paths labeled by $x$ lead to a final state; the existence of one such path suffices to put $x$ into the language $L(M)$.  

Note that $ab \in L(M)$ because of the existence of the path $(q_0, q_1, q_3)$ labeled by this word and the fact that $q_3$ is a final state. On the other hand, $(q_0, q_2, q_4)$ is another path labeled by $x$ but $q_4 \notin F$. 
Example (cont’d)

This ndfa is simple enough to allow an easy identification of all types of words in $L(\mathcal{M})$:

1. The final state $q_1$ can be reached by applying an arbitrary number of $b$s followed by an $a$.

2. The final state $q_3$ can be reached by a path of the form $(q_0, \ldots, q_0, q_1, q_3)$, that is by a word of the form $b^kab$ for $k \in \mathbb{N}$.

3. The same final state $q_3$ can be reached via $q_2$. Input words that allow this transition have the form $b^kaa$ for $k \in \mathbb{N}$.

Thus, we have

$$L(\mathcal{M}) = \{b\}^*a \cup \{b\}^*ab \cup \{b\}^*aa = \{b\}^*\{a, ab, aa\}.$$
Example

Consider an alphabet \( A = \{a_0, \ldots, a_{n-1}\} \) and a binary relation \( \rho \subseteq A \times A \). The language

\[
L_\rho = \{ a_{i_0} \cdots a_{i_p} \mid p \in \mathbb{N}, (a_{i_j}, a_{i_{j+1}}) \in \rho \text{ for } 0 \leq j \leq p - 1 \}
\]

is accepted by the ndfa \( M_\rho = (A, Q, \delta, q, F) \), where \( Q = \{q, q_0, \ldots, q_{n-1}\} \), \( F = \{q_0, \ldots, q_{n-1}\} \), and \( \delta \) is given by

- \( \delta(q, a_i) = \{q_i\} \) for \( 0 \leq i \leq n - 1 \);
- for every \( i, j \) such that \( 0 \leq i, j \leq n - 1 \),

\[
\delta(q_i, a_j) = \{q_j \in Q \mid (a_i, a_j) \in \rho\}.
\]
Example (cont’d)

Note that if \((a_i, a_j) \notin \rho\), then \((q_i, a_j)\) is a blocking situation. The existence of these blocking situations is precisely what makes this device a nondeterministic automaton.

Graph of the ndfa

\[
\begin{align*}
q & \quad a_0 \quad \vdots \\
q & \quad a_i \quad q_i \\
q & \quad a_j \\
q & \quad a_{n-1} \quad \vdots \\
q & \quad q_{n-1}
\end{align*}
\]
We show that $L_\rho = L(M_\rho)$. Let $x = a_{i_0} \cdots a_{i_p}$ be a word in $L_\rho$ with $p \geq 0$. We prove by induction on $p = |x| - 1$ that $x \in L(M_\rho)$ and that $\delta^*(q, x) = \{q_{i_p}\}$. The base case, $p = 0$, is immediate, since the condition $(a_{i_j}, a_{i_{j+1}}) \in \rho$ for $0 \leq j \leq p - 1$ is vacuous.

Suppose that the statement holds for words in $L_\rho$ of length at most $p$ and let $x = a_{i_0} \cdots a_{i_p}$ be a word in $L_\rho$ of length $p + 1$. It is clear that the word $y = a_{i_0} \cdots a_{i_{p-1}}$ belongs to $L_\rho$. By the inductive hypothesis, $y \in L(M_\rho)$ and $\delta^*(q, y) = \{q_{i_{p-1}}\}$. Since $(a_{i_{p-1}}, a_{i_p}) \in \rho$ (by the definition of $L_\rho$), we have $\delta(q_{i_{p-1}}, a_{i_p}) = \{q_{i_p}\}$, so

$$q_{i_p} \in \bigcup_{q' \in \delta(q, y)} \delta(q', a_{i_p}) = \delta^*(q, ya_{i_p}) = \delta^*(q, x).$$

Therefore, $x \in L(M_\rho)$.  


To prove the converse inclusion $L(M_\rho) \subseteq L_\rho$ we use an argument by induction on $|x| \geq 1$, where $x$ is a word from $L(M_\rho)$, to show that if $x = a_{i_0} \cdots a_{i_p} \in L(M_\rho)$, then $\delta^*(q, x) = \{q_{i_p}\}$ and $x \in L_\rho$. Again, the base case is immediate.

Suppose that the statement holds for words in $L(M_\rho)$ of length less than $p + 1$ that belong to $L_\rho$, and let $x = a_{i_0} \cdots a_{i_p}$ be a word in $L(M_\rho)$ of length $p + 1$. If $y = a_{i_0} \cdots a_{i_{p-1}}$, it is easy to see that $y \in L(M_\rho)$ because no blocking situation may arise in $M_\rho$ while the symbols of $y$ are read. Therefore, by the inductive hypothesis, $y \in L_\rho$ and $\delta^*(q, y) = \{q_{i_{p-1}}\}$.

Further, since $\delta(q_{i_{p-1}}, a_{i_p}) \neq \emptyset$, it follows that $(a_{i_{p-1}}, a_{i_p}) \in \rho$, so $x \in L_\rho$, and $\delta^*(q, x) = q_{i_p}$.

Thus, $L_\rho$ is accepted by the ndfa $M_\rho$. 


Let $M = (A, Q, \delta, q_0, F)$ be a nondeterministic automaton, and let 
$\Delta : \mathcal{P}(Q) \times A \rightarrow \mathcal{P}(Q)$ be defined by

$$\Delta(S, a) = \bigcup_{q \in S} \delta(q, a)$$

(1)

for every $S \subseteq Q$ and $a \in A$. Starting from $\Delta$, we define 
$\Delta^* : \mathcal{P}(Q) \times A^* \rightarrow \mathcal{P}(Q)$ in the manner used for the transition functions 
of deterministic automata. Namely, we define

$$\Delta^*(S, \lambda) = S$$

(2)

$$\Delta^*(S, xa) = \Delta(\Delta^*(S, x), a)$$

(3)

for every $S \subseteq Q$ and $a \in A$. 
Lemma

The functions $\Delta$ and $\Delta^*$ defined above satisfy the following properties:

1. For every family of sets $\{S_0, \ldots, S_{n-1}\}$ and every $a \in A$, we have:

$$\Delta \left( \bigcup_{0 \leq i \leq n-1} S_i, a \right) = \bigcup_{0 \leq i \leq n-1} \Delta(S_i, a).$$

2. For every set $S \subseteq Q$ and $x \in A^*$ we have

$$\Delta^*(S, x) = \bigcup_{q \in S} \delta^*(q, x).$$
Proof

The first part of the lemma is immediate, because

\[
\Delta(\bigcup_{0 \leq i \leq n-1} S_i, a) = \bigcup\{\delta(q, a) \mid q \in \bigcup_{0 \leq i \leq n-1} S_i\}
= \bigcup_{0 \leq i \leq n-1}\{\delta(q, a) \mid q \in S_i\}
= \bigcup_{0 \leq i \leq n-1}\Delta(S_i, a).
\]
The argument for the second part of the lemma is by induction on $|x|$. For the basis step, we have $|x| = 0$, so $x = \lambda$, and $\Delta^*(S, \lambda) = S$, $\bigcup_{q \in S} \delta^*(q, \lambda) = \bigcup_{q \in S} \{q\} = S$.

Suppose that the argument holds for words of length $n$, and let $x = za$ be a word of length $n + 1$. We have

\[
\Delta^*(S, x) = \Delta^*(S, za) = \Delta(\Delta^*(S, z), a) = \Delta(\bigcup_{q \in S} \delta^*(q, z), a) \tag{by ind. hyp.}
\]

\[
= \bigcup_{q \in S} \Delta(\delta^*(q, z), a)
\]

\[
= \bigcup_{q \in S} \delta(r, a) = \bigcup_{q \in S} \delta^*(q, za) = \bigcup_{q \in S} \delta^*(q, x).
\]
Nondeterministic automata can be regarded as generalizations of deterministic automata in the following sense. If $M = (A, Q, \delta, q_0, F)$ is a deterministic automaton, consider a nondeterministic automaton $M' = (A, Q, \delta', q_0, F)$, where $\delta'(q, a) = \{\delta(q, a)\}$. It is easy to verify that for every $q \in Q$ and $x \in A^*$ we have $\delta'^*(q, x) = \{\delta^*(q, x)\}$. Therefore,

$$L(M') = \{x \in A^* \mid \delta'^*(q_0, x) \cap F \neq \emptyset\}$$

$$= \{x \in A^* \mid \{\delta^*(q_0, x)\} \cap F \neq \emptyset\}$$

$$= \{x \in A^* \mid \delta^*(q_0, x) \in F\}$$

$$= L(M).$$

In other words, for every deterministic finite automaton there exists a nondeterministic one that recognizes the same language.
Theorem

For every nondeterministic automaton, there exists a deterministic automaton that accepts the same language.
Proof

Let \( M = (A, Q, \delta, q_0, F) \) be a nondeterministic automaton. Define the function \( \Delta \) as in the equality

\[
\Delta(S, a) = \bigcup_{q \in S} \delta(q, a),
\]

and consider the deterministic automaton

\( M' = (A, \mathcal{P}(Q), \Delta, \{q_0\}, \{S \mid S \subseteq Q \text{ and } S \cap F \neq \emptyset\}) \). For every \( x \in A^* \) we have the following equivalent statements:

1. \( x \in L(M) \);
2. \( \delta^*(q_0, x) \cap F \neq \emptyset \);
3. \( \Delta^*(\{q_0\}, x) \cap F \neq \emptyset \);
4. \( x \in L(M') \).

This proves that \( L(M) = L(M') \).
Example
Consider the nondeterministic finite automaton

\[ M = (\{a, b\}, \{q_0, q_1, q_2\}, \delta, q_0, \{q_2\}) \]

whose graph is given below.
It is easy to see that the language accepted by this automaton is $A^*bbA^*$, that is the language that consists of all words that contain two consecutive $b$ symbols.

The graph of the nondeterministic automaton is simpler than the graph of the previous deterministic automaton; this simplification is made possible by the nondeterminism.
Graph of the Equivalent ndfa