Finite Automata and Regular Languages (part III)

Prof. Dan A. Simovici

UMB

1. Transition Systems

Nondeterministic finite automata can be further generalized by allowing transitions between states without reading any input symbol.

Definition

A transition system (ts) is a 5-tuple $\mathcal{T} = (A, Q, \theta, Q_0, F)$, where A, Q, and F are as in a finite automaton, θ is a finite relation, $\theta \subseteq Q \times A^* \times Q$, called the transition relation of $\mathcal{T}$, and Q_0 is a nonempty subset of Q called the set of initial states, and F is the set of final states.
A transition in $\mathcal{T}$ is a triple $(q, x, q') \in \theta$. We refer to transitions of the form (q, λ, q') as null transitions.

Transition systems are conveniently represented by labeled directed multigraphs. Namely, if $\mathcal{T} = (A, Q, \theta, Q_0, F)$ is a transition system, then its graph is a labeled directed multigraph $G(\mathcal{T})$.

- $G(\mathcal{T})$ has Q as its set of vertices;
- each directed edge e from q to q' labelled x corresponds to a triple $(q, x, q') \in \theta$, and every such triple is represented by an edge in G.

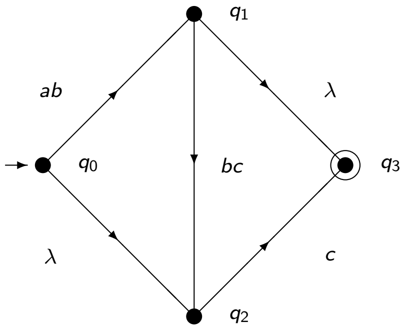
Unlike the graph of a dfa or an ndfa, the edges of the directed graph of a transition system can be labelled by words, including the null word.

Example

The graph of the transition system

$$\mathcal{T}_1 = (\{a, b, c\}, \{q_0, q_1, q_2, q_3\}, \theta, \{q_0\}, \{q_3\}),$$

where θ is given by
$\theta = \{(q_0, ab, q_1), (q_0, \lambda, q_2), (q_1, bc, q_2), (q_1, \lambda, q_3), (q_2, c, q_3)\}$ is shown below:



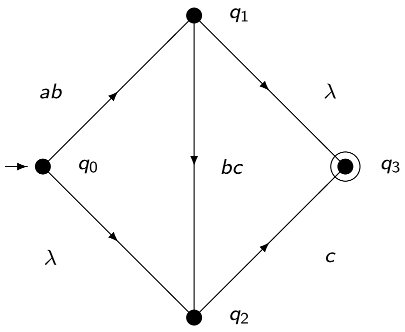
Extending the transition relation

As we did with the transition functions of dfas and ndfas, we wish to extend the transition relation θ of a transition system $\mathcal{T}$ to the set $Q \times A^* \times Q$. The extension $\theta^* \subseteq Q \times A^* \times Q$ is given next.

1. For every $q \in Q$ define $(q, \lambda, q) \in \theta^*$.
2. Every triple $(q, x, q') \in \theta$ belongs to θ^*.
3. If $(q, x, q'), (q', y, q'') \in \theta^*$, then $(q, xy, q'') \in \theta^*$.

Note that $(q, w, q') \in \theta^*$ if and only if there is a path in $G(\mathcal{T})$ that begins with q and ends with q' such that the concatenated labels of the directed edges of this path form the word w.

Example



If $\mathcal{T}$ is the above transition system, then $(q_0, abbcc, q_3) \in \theta^*$ because $(q_0, ab, q_1), (q_1, bc, q_2), (q_2, c, q_3) \in \theta$. Similarly, $(q_0, c, q_3) \in \theta^*$ because $(q_0, \lambda, q_2), (q_2, c, q_3) \in \theta$.

Definition

Let $\mathcal{T} = (A, Q, \theta, Q_0, F)$ be a transition system. The language accepted by $\mathcal{T}$ is

$$L(\mathcal{T}) = \{x \in A^* \mid (q_0, x, q) \in \theta^* \text{ for some } q_0 \in Q_0, q \in F\}.$$

Thus, a word x belongs to $L(\mathcal{T})$ if there is a path in $G(\mathcal{T})$ that begins in an initial state $q_0 \in Q_0$, labeled by x, such that the path ends in one of the states of F, the set of final states.

Transition systems generalize ndfas

If $\mathcal{M} = (A, Q, \delta, q_0, F)$ is a nondeterministic automaton, define the transition system $\mathcal{T}_{\mathcal{M}} = (A, Q, \theta, \{q_0\}, F)$, where

$$\theta = \{(q, a, q') \mid q, q' \in Q, a \in A \text{ and } q' \in \delta(q, a)\}.$$

It can be shown by induction on $|x|$ that $(q, x, q') \in \theta^*$ if and only if $q' \in \delta^*(q, x)$. This implies that $L(\mathcal{T}_{\mathcal{M}}) = L(\mathcal{M})$. Therefore, every regular language can be accepted by a transition system. Furthermore, any language that can be accepted by a transition system is regular.

Lemma

For every transition system $\mathcal{T} = (A, Q, \theta, Q_0, F)$ *there exists a transition system* $\mathcal{T}' = (A, Q', \theta', Q_0, F)$ *such that* $(q, x, q_1) \in \theta'$ *implies* $|x| \leq 1$ *and* $L(\mathcal{T}') = L(\mathcal{T})$.

Proof

Define the relation θ' and the set Q' as follows:

1. Every state $q \in Q$ also belongs to Q'.
2. Every triple $(q, x, q') \in \theta$ such that $|x| \leq 1$ also belongs to θ'.
3. If $t = (q, x, q') \in \theta$ such that $x = a_0 \ldots a_{n-1}$ and $n \geq 2$, add $n-1$ new states $q_0^t, \ldots q_{n-2}^t$ to Q' and the triples

$$(q, a_0, q_0^t), (q_0^t, a_1, q_1^t), \ldots, (q_{n-2}^t, a_{n-1}, q')$$

   to θ'.

The ts $\mathcal{T}'$ clearly satisfies the conditions of the lemma, since $(q, x, q') \in \theta^*$ if and only if $(q, x, q') \in \theta'^*$.

Theorem

For every transition system $\mathcal{T} = (A, Q, \theta, Q_0, F)$ *there exists a deterministic finite automaton* $\mathcal{M}$ *such that* $L(\mathcal{T}) = L(\mathcal{M})$.

Proof.

By the previous Lemma we can assume that $(q, x, q') \in \theta$ implies $|x| \leq 1$. Define the deterministic finite automaton $\mathcal{M} = (A, \mathcal{P}(Q), \Delta, Q_0', F')$, where the initial state of $\mathcal{M}$ is

$$Q_0' = \{q \in Q \mid (q_0, \lambda, q) \in \theta^* \text{ for some } q_0 \in Q_0\},$$

the set of final states is $F' = \{S \mid S \subseteq Q, S \cap F \neq \emptyset\}$, and the function Δ is defined by

$$\Delta(S, a) = \{q' \in Q \mid (q, a, q') \in \theta^* \text{ for some } q \in S\},$$

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

Proof (cont'd)

It is not difficult to verify, by induction on $|x|$, that

$$\Delta^*(Q_0', x) = \{q' \in Q \mid (q_0, x, q') \in \theta^* \text{ for some } q_0 \in Q_0\},$$

for $x \in A^*$. For the basis case, $|x| = 0$, so the above equality becomes

$$Q_0' = \{q' \in Q \mid (q_0, \lambda, q') \in \theta^* \text{ for some } q_0 \in Q_0\},$$

which holds by the definition of Q_0'.

Proof (cont'd)

Suppose that the equality holds for words of length n, and let y be a word of length $n+1$. We can write $y = xa$, so

$$
\begin{aligned}
\Delta^*(Q_0', y) &= \Delta^*(Q_0', xa) \\
&= \Delta(\Delta^*(Q_0', x), a) \\
&= \Delta(\{q' \in Q \mid (q_0, x, q') \in \theta^* \text{ for some } q_0 \in Q_0\}, a) \\
&\quad \text{(by the inductive hypothesis)} \\
&= \{r \in Q \mid (q', a, r) \in \theta^*, \text{ for some } q' \text{ such that} \\
&\quad (q_0, x, q') \in \theta^* \text{ for some } q_0 \in Q_0\} \\
&\quad \text{(by the definition of } \Delta) \\
&= \{r \in Q \mid (q_0, xa, r) \in \theta^* \text{ for some } q_0 \in Q_0\} \\
&\quad \text{(by the definition of } \theta^*) \\
&= \{r \in Q \mid (q_0, y, r) \in \theta^* \text{ for some } q_0 \in Q_0\},
\end{aligned}
$$

which concludes our inductive argument.

Proof (cont'd)

From this it follows that $L(\mathcal{M}) = L(\mathcal{T})$. By definition, $x \in L(\mathcal{M})$ if and only if $\Delta^*(Q_0', x) \in F'$. This is equivalent to $\Delta^*(Q_0', x) \cap F \neq \emptyset$. This is equivalent to the existence of a state $q' \in F'$ such that $(q_0, x, q') \in \theta^*$ for some $q_0 \in Q_0$, and this is equivalent to $x \in L(\mathcal{T})$.

Corollary

The class of languages that are accepted by transition systems is the class $\mathcal{R}$ of regular languages.

Definition

Let $\mathcal{T} = (A, Q, \theta, Q_0, F)$ be a transition system. The λ-closure is the mapping $K_{\mathcal{T}} : \mathcal{P}(Q) \longrightarrow \mathcal{P}(Q)$ given by

$$K_{\mathcal{T}}(S) = \{q \in Q \mid (s, \lambda, q) \in \theta^* \text{ for some } s \in S\}.$$

The set $K_{\mathcal{T}}(S)$ comprises the states in S plus all the states that can be reached from a state in S using a series of λ-transitions.

Theorem

Let $\mathcal{T} = (A, Q, \theta, Q_0, F)$ be a transition system. The λ-closure of $\mathcal{T}$ has the following properties.

1. $S \subseteq K_{\mathcal{T}}(S)$;
2. $S \subseteq S'$ *implies* $K_{\mathcal{T}}(S) \subseteq K_{\mathcal{T}}(S')$;
3. $K_{\mathcal{T}}(K_{\mathcal{T}}(S)) = K_{\mathcal{T}}(S)$,

for every $S, S' \in \mathcal{P}(Q)$.

Proof

Since $(s, \lambda, s) \in \theta^*$ it is immediate that $S \subseteq K_{\mathcal{T}}(S)$ for every $S \in \mathcal{P}(Q)$. The second part of the theorem is a direct consequence of the definition of $K_{\mathcal{T}}$.
Note that Parts (i) and (ii) imply $K_{\mathcal{T}}(S) \subseteq K_{\mathcal{T}}(K_{\mathcal{T}}(S))$. Let $q \in K_{\mathcal{T}}(K_{\mathcal{T}}(S))$. There is a state $s \in S$ and a state $r \in K_{\mathcal{T}}(S)$ such that $(s, \lambda, r) \in \theta^*$ and $(r, \lambda, q) \in \theta^*$. By the definition of θ^* we obtain $(s, \lambda, q) \in \theta^*$, so $q \in K_{\mathcal{T}}(S)$. This implies $K_{\mathcal{T}}(K_{\mathcal{T}}(S)) \subseteq K_{\mathcal{T}}(S)$, which gives the last part of the theorem.

Definition

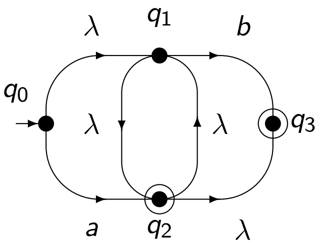
Let $\mathcal{T} = (A, Q, \theta, Q_0, F)$ be a transition system. A $K_{\mathcal{T}}$-closed subset of Q is a set S such that $S \subseteq Q$ and $K_{\mathcal{T}}(S) = S$.

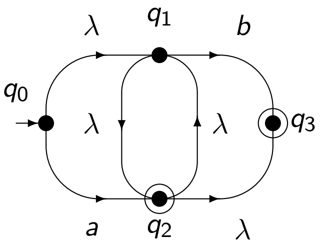
Example

Consider the transition system

$$\mathcal{T} = (\{a, b\}, \{q_0, q_1, q_2, q_3\}, \theta, \{q_0\}, \{q_2, q_3\})$$

whose graph is shown





S	$K_{\mathcal{T}}(S)$	S	$K_{\mathcal{T}}(S)$
$\emptyset$	$\emptyset$	$\{q_1, q_2\}$	$\{q_1, q_2, q_3\}$
$\{q_0\}$	Q	$\{q_1, q_3\}$	$\{q_1, q_2, q_3\}$
$\{q_1\}$	$\{q_1, q_2, q_3\}$	$\{q_2, q_3\}$	$\{q_1, q_2, q_3\}$
$\{q_2\}$	$\{q_1, q_2, q_3\}$	$\{q_0, q_1, q_2\}$	Q
$\{q_3\}$	$\{q_3\}$	$\{q_0, q_1, q_3\}$	Q
$\{q_0, q_1\}$	Q	$\{q_0, q_2, q_3\}$	Q
$\{q_0, q_2\}$	Q	$\{q_1, q_2, q_3\}$	$\{q_1, q_2, q_3\}$
$\{q_0, q_3\}$	Q	$\{q_0, q_1, q_2, q_3\}$	Q

The closed subsets of Q are $\emptyset$, $\{q_3\}$, $\{q_1, q_2, q_3\}$, and Q itself.

Theorem

Let $\mathcal{T} = (A, Q, \theta, Q_0, F)$ *be a transition system and let* $\mathcal{M} = (A, \mathcal{P}(Q), \Delta, Q_0', F')$ *be the dfa constructed in earlier. The set* $\Delta(S, a)$ *is a* $K_{\mathcal{T}}$*-closed set of states for every subset* S *of* Q *and* $a \in A$.

Proof

To prove the theorem it suffices to show that $K_{\mathfrak{T}}(\Delta(S, a)) \subseteq \Delta(S, a)$. Let $p \in K_{\mathfrak{T}}(\Delta(S, a))$. There is $p_1 \in \Delta(S, a)$ such that $(p_1, \lambda, p) \in \theta^*$. The definition of $\Delta(S, a)$ implies the existence of $q \in S$ such that $(q, a, p_1) \in \theta^*$. Thus, $(q, a, p) \in \theta^*$, so $p \in \Delta(q, a)$.

Corollary

Let $\mathcal{T} = (A, Q, \theta, Q_0, F)$ *be a transition system, and let* $\mathcal{M} = (A, \mathcal{P}(Q), \Delta, Q_0', F')$ *be the constructed dfa. The accessible states of the dfa* $\mathcal{M}$ *are* $K_{\mathcal{T}}$*-closed subsets of* Q*.*

Proof.

The initial state Q_0' of $\mathcal{M}$ is obviously closed. If Q' is an accessible state of $\mathcal{M}$, then $Q' = \Delta(S, a)$ for some $S \subseteq Q$. Therefore Q' is closed. □

Theorem

Let $\mathcal{T} = (A, Q, \theta, Q_0, F)$ *be a transition system, and let* $\mathcal{M} = (A, \mathcal{P}(Q), \Delta, Q_0', F')$ *be the dfa constructed earlier. Then,* $\Delta(S, a) = K_{\mathcal{T}}(\{q \in Q \mid (s, a, q) \in \theta \text{ for some } s \in S\})$*, where* S *is an accessible state of* $\mathcal{M}$*.*

Proof

Let $s \in S$. Note that if $(s, a, q) \in \theta$ and $(q, \lambda, q_1) \in \theta^*$, then by the definition of θ^* we have $(s, a, q_1) \in \theta^*$. Therefore, $K_{\mathfrak{T}}(\{q \in Q \mid (s, a, q) \in \theta\} \subseteq \Delta(S, a)$.
To prove the converse inclusion, $\Delta(S, a) \subseteq K_{\mathfrak{T}}(\{q \in Q \mid (s, a, q) \in \theta\}$, let $(s, a, q_1) \in \theta^*$. Then, there is a path in $G(\mathfrak{T})$ that begins with s and ends with q_1 such that the concatenated labels of the directed edges of this path form the word a. This implies the existence of the states $p, p' \in Q$ such that $(s, \lambda, p) \in \theta^*$, $(p, a, p_1) \in \theta$, and $(p_1, \lambda, q_1) \in \theta^*$. Since S is $K_{\mathfrak{T}}$-closed it follows that $p \in S$ and this gives the desired conclusion.

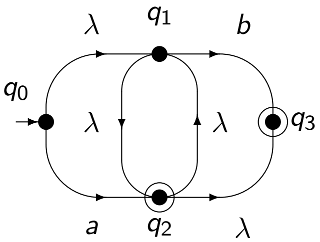
An Algorithm for Constructing a dfa corresponding to a ts

Input: A transition system $\mathcal{T} = (A, Q, \theta, Q_0, F)$.
Output: An accessible dfa $\mathcal{M}_1$ such that $L(\mathcal{M}_1) = L(\mathcal{T})$.
Method: Compute the increasing sequence of collections of subsets of Q, $\mathcal{Q}_0, \ldots, \mathcal{Q}_i, \ldots$, where

$$\begin{aligned} \mathcal{Q}_0 &= \{Q_0'\} \\ \mathcal{Q}_{i+1} &= \mathcal{Q}_i \cup \{U \in \mathcal{P}(Q) \mid U = \Delta(S, a) \text{ for some } S \in \mathcal{Q}_i \text{ and } a \in A\}. \end{aligned}$$

the computation of $U = \Delta(S, a)$ can be done by computing first the set $W = \{q \in Q \mid (s, a, q) \in \theta \text{ for some } s \in S\}$ and then $U = K_{\mathcal{T}}(W)$. Stop when $\mathcal{Q}_{i+1} = \mathcal{Q}_i$. The set $\mathcal{Q}_i$ is the set of accessible states of $\mathcal{M}$. Output $\mathcal{M}' = \mathrm{ACC}(\mathcal{M})$, the accessible component of $\mathcal{M}$.

For the transition system



the transition system is: