DFA with a Bounded Activity Level

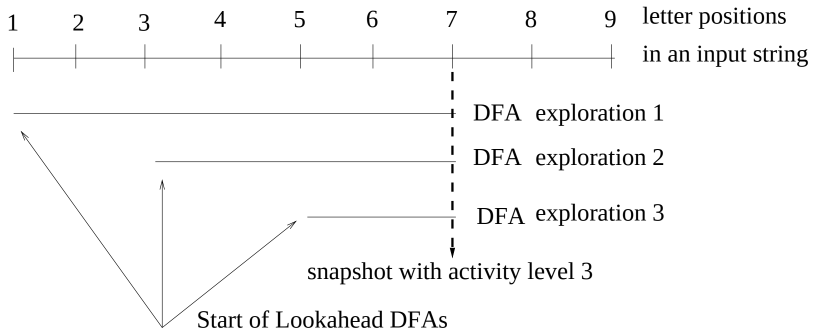
Marius Konitzer and Hans U. Simon
hans.simon@rub.de

Department of Mathematics
Ruhr-University Bochum

8th Int. Conf. on Language and Automata Theory and Applications

Outline

LR-regular Parsing and Unbounded Lookahead



LR-regular Parsing and Time Bounds

- Theoretically, LR-regular Parsing can be done in linear time.
  Čulik and Cohen, “LR-regular Grammars — An Extension of LR(k) Grammars”, JCSS 7(1), 1973.
- Most practical LR-regular parsers have the following time bounds:

$$
\begin{array}{lcl}
n & = & \text{length of the input string} \\
\ell & = & \text{number of simultaneous DFA-explorations} \\
 & \Longrightarrow & \text{run-time } \; O(\ell n) = \begin{cases} O(n) & \text{in the best case} \\ O(n^2) & \text{in the worst case} \end{cases}
\end{array}
$$

Bounded versus Unbounded Avtivity Level of a DFA

Consider: DFA M with a partially defined transition function δ.

Interpretation: $\delta(q, a)$ undefined $\iff$ computation stops

Definition

We say a DFA M has an activity level of at least ℓ if there exists a string $w \in \Sigma^n$ and $1 \leq i_1 < \ldots < i_\ell < j \leq n$ such that the DFA-computations started at letter positions $i_1, \ldots, i_\ell$ are still active at letter position j.

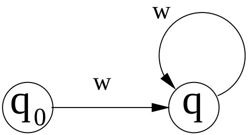
ℓ_M = largest possible activity level of M (possibly ∞)

Prefix-cyclic DFA

Definition

M prefix-cyclic $\iff \exists$ state q, $\exists$ string w:
$\delta^*(q_0, w) = q = \delta^*(q, w)$.

In the transition diagram of M:



Prefix-cyclic = Unbounded Activity Level

Theorem

M is prefix-cyclic $\iff \ell_M = \infty$.

$\implies$: let $w = w_1 \cdots w_k$ such that $\delta^*(q_0, w) = q = \delta^*(q, w)$.

w_1 … w_k w_1 … w_k w_1 … w_k

q_o

q, q_o

q, q, q_o

q exploration 1

q exploration 2

q exploration 3

q_o exploration 4

and so on

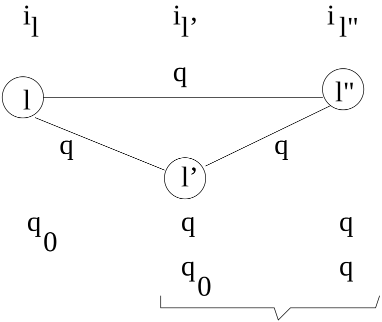
Thus: $\ell_M = \infty$.

We show: $\ell_M = \infty \implies M$ is prefix-cyclic.

- Let K_ℓ denote the complete graph with ℓ nodes.
- Let $r(3,t)$ be the smallest ℓ such that every coloring of the edges of K_ℓ induces a monochromatic triangle (triangular Ramsey number with t colors).
- From now let $t := |Q|$ (the number of states of M).
- Let $\ell = r(3,t)$ and let $w \in \Sigma^n$ be a word with letter positions $1 \leq i_1 < i_2 < \ldots < i_\ell < n$ witnessing that M has activity level at least ℓ.
- Let $1 \leq l < l' \leq \ell$. Color the edge $\{l, l'\}$ by the state of M reached in letter position l' when M starts its computation in letter position l.

We claim that "monochromatic triangle in K_ℓ" translates to "M being prefix-cyclic".

Proof of the Claim:



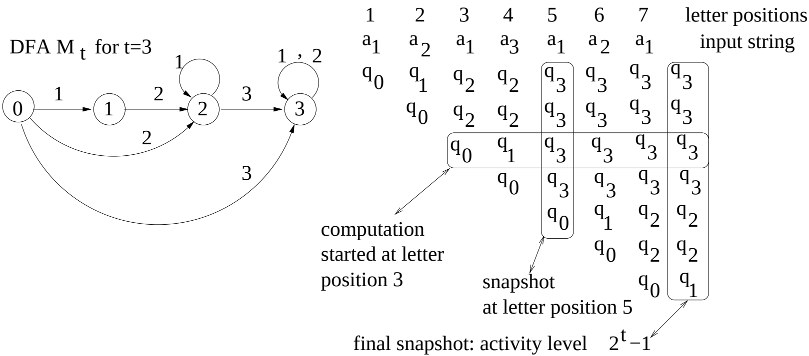
letter positions in w

monochromatic triangle
in the complete graph

corresponding state transitions
(alligned with the letter positions)

The subword of w between letter positions $i_{l'}$ and $i_{l''}$
witnesses that M is prefix−cyclic.

DFA with 3 non-initial states and activity level 7



The input string for the “next” DFA M_4 would be:

$$a_1a_2a_1a_3a_1a_2a_1 \;\; a_4 \;\; a_1a_2a_1a_3a_1a_2a_1 \;.$$

DFA with t non-initial states and activity level $2^t - 1$

Let $M_t = (Q, \Sigma, \delta, q_0)$ be given by $Q = \{q_0, q_1, \dots, q_t\}$, $\Sigma = \{a_1, \dots, a_t\}$, and

$$\delta(q_i, a_j) = \begin{cases} q_j & \text{if } i < j \\ q_i & \text{if } i > j \\ \text{undefined} & \text{if } i = j \end{cases} . \qquad (1)$$

Central Observation: If $w(t-1) \in \{a_1, \dots, a_{t-1}\}^*$ witnesses that M_{t-1} has activity level at least $2^{t-1} - 1$, then $w(t) = w(t-1)a_t w(t-1)$ witnesses that M_t has activity level at least $2^t - 1$.

Conclusion: For every t there exists a DFA with t non-initial states, an input alphabet of size t and activity level (at least) $2^t - 1$.

Alphabet Size Does Not Matter Much

Theorem

For every t *there exists a DFA with* $1 + 3t\lceil\log t\rceil$ *states, a* binary input alphabet *and activity level* $2^t - 1$*.*

A Tight Upper Bound

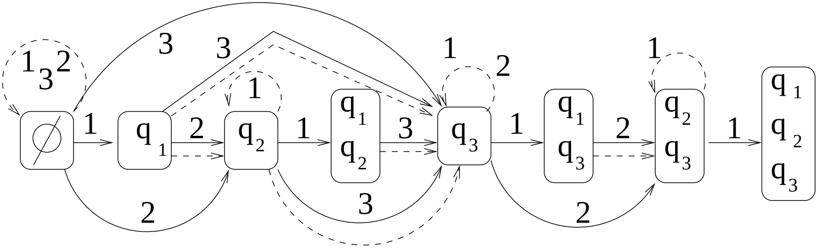
Theorem

M is a not prefix-cyclic DFA with t non-initial states
$\Longrightarrow \ell_M \leq 2^t - 1$.

The "snapshot graph" G is a digraph whose nodes are the snapshots and whose arcs (solid or dotted) are as follows:

- Let $a \in \Sigma$. If $\delta(q, a)$ is defined for all $q \in S$, we draw a dotted arc from S to $S'(a) = \{\delta(q, a) | \; q \in S\}$.
- If $\delta(q_0, a)$ is defined too, then draw a solid arc from S to $S'' = S' \cup \{\delta(q_0, a)\}$.

Example of a Snapshot Graph



For a snapshot graph G (with at most 2^t nodes) of a DFA M with t non-initial states) the following holds:

- The paths through G describe the evolution of snapshots induced by the currently active computations of M.
- When passing through a solid arc, the current activity level increases by 1.
- Every cycle in G consists of dotted arcs only. (Otherwise the activity level could be pushed to ∞.)
- If we assign length 1 to solid arcs and length 0 to dotted ones, then ℓ_M coincides with the total length of a longest **simple** path in G.
- Thus, $\ell_M \leq 2^t - 1$, as claimed.

Testing Prefix-cyclicity

Theorem

It can be checked in quadratic time whether a given DFA is prefix-cyclic.

Proof.

Build the product automaton M^2 of M with itself and check whether in the transition diagram of M^2 there is a path from (q_0, q) to (q, q). □

Computing the Maximum Activity Level

Theorem

Given a not prefix-cyclic DFA M, ℓ_M can be computed in time linear in the size of the corresponding snapshot graph G_M.

Proof.

Before computing the longest path, replace G_M by the **acyclic** supergraph which represents each strongly connected component of G_M by a single supernode. □

Theorem

Given a DFA M and a threshold T, the problem to decide whether $\ell_M > T$ is PSPACE-complete.

Proof: Polynomial reduction from "Finite Automata Intersection"

Summary

- The activity levels of Lookahead DFA influence the time bounds of LRR-parsing techniques.
- A DFA has an unbounded activity level iff it is prefix-cyclic — a property that can be checked in quadratic time.
- A not prefix-cyclic DFA with t non-initial states has an activity level of at most $2^t - 1$ and this bound is tight.
- The computation of the maximum activity level of a given DFA is PSPACE-hard (and PSPACE-easy), but there is an algorithm whose run-time is linear in the size of the corresponding snapshot graph.

Thank you . . .

. . . for your attention!