Available online at www.sciencedirect.com
SCIENCE DIRECT®

Information Processing Letters 98 (2006) 231–234

Information Processing Letters

www.elsevier.com/locate/ipl

The state complexity of L^2 and L^k

Narad Rampersad

School of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Received 21 October 2004; received in revised form 17 June 2005; accepted 17 June 2005

Available online 20 March 2006

Communicated by L. Boasson

Abstract

We show that if M is a DFA with n states over an alphabet with at least two letters and $L = L(M)$, then the worst-case state complexity of L^2 is $n2^n - 2^{n-1}$. If, however, M is a DFA over a unary alphabet, then the worst-case state complexity of L^k is $kn - k + 1$ for all $k \geqslant 2$.

Keywords: Formal languages; State complexity

1. Introduction

We are often interested in quantifying the complexity of a regular language L. One natural complexity measure for regular languages is the *state complexity* of L, that is, the number of states in the minimal deterministic finite automaton (DFA) that accepts L. Given an operation on regular languages, we may also define the state complexity of that operation to be the number of states that are both sufficient and necessary in the worst-case for a DFA to accept the resulting language.

Birget [1] gave exact results for the state complexities of the intersection and union operations on regular languages. Yu et al. [10] studied other operations, such as concatenation and Kleene star. For instance, Yu et al. proved that, given DFAs M_1 and M_2 with m and n states, respectively, there exists a DFA with $m2^n - 2^{n-1}$ states that accepts $L(M_1)L(M_2)$. Moreover, there exist M_1 and M_2 for which this bound is optimal. Some more recent work on the state complexity of concatenation has been done by Jirásková [5] as well as Jirásek et al. [6]. Birget's work [2] on the state complexity of $\overline{\Sigma^*\overline{L}}$ may also be of interest.

We are interested here in the state complexity of the concatenation of a regular language L with itself, which we denote L^2. We show that the bounds of Yu et al. for concatenation are also optimal for L^2. In other words, if M is a DFA with n states and $L = L(M)$, then the worst-case state complexity of L^2 is $n2^n - 2^{n-1}$. This bound, however, does not hold if we restrict ourselves to unary languages. Specifically, we show that if M is a DFA over a unary alphabet, then the worst-case state complexity of L^k is $kn - k + 1$ for all $k \geqslant 2$.

We first recall some basic definitions. For further details see [4]. A *deterministic finite automaton* M is a quintuple $M = (Q, \Sigma, \delta, q_0, F)$, where Q is a finite set of states; Σ is a finite alphabet; $\delta : Q \times \Sigma \to Q$ is the transition function, which we extend to $Q \times \Sigma^*$ in the natural way; $q_0 \in Q$ is the start state; and $F \subseteq Q$ is the set of final states. A DFA M accepts a word $w \in \Sigma^*$ if

E-mail address: nrampersad@cs.uwaterloo.ca (N. Rampersad).

0020-0190/$ – see front matter
doi:10.1016/j.ipl.2005.06.011

$\delta(q_0, w) \in F$. The language accepted by M is the set of all $w \in \Sigma^*$ such that $\delta(q_0, w) \in F$; this language is denoted $L(M)$. We denote the language $L(M)L(M)$ by $L^2(M)$. We may extend this notation to higher powers by the recursive definition $L^k(M) = L^{k-1}(M)L(M)$ for $k \geqslant 2$.

2. State complexity of L^2 for binary alphabets

In this section we consider the state complexity of L^2 for languages L over an alphabet of size at least 2.

Theorem 1. *For any integer $n \geqslant 3$, there exists a DFA M with n states such that the minimal DFA accepting the language $L^2(M)$ has $n2^n - 2^{n-1}$ states.*

Proof. That the minimal DFA for $L^2(M)$ has at most $n2^n - 2^{n-1}$ states follows from the upper bound of Yu et al. for concatenation of regular languages mentioned in the introduction. To show that $n2^n - 2^{n-1}$ states are also necessary in the worst case we define a DFA $M = (Q, \Sigma, \delta, 0, F)$ (Fig. 1), where $Q = \{0, \dots, n-1\}$, $\Sigma = \{0, 1\}$, $F = \{n-1\}$, and for any i, $0 \leqslant i \leqslant n-1$,

$$\delta(i, a) = \begin{cases} 0 & \text{if } a = 0 \text{ and } i = 1, \\ i & \text{if } a = 0 \text{ and } i \neq 1, \\ i + 1 \bmod n & \text{if } a = 1. \end{cases}$$

We will apply the construction of Yu et al. [10, Theorem 2.3] and show that the resulting DFA for $L^2(M)$ is minimal (see [6] for another example of this approach). Let $M' = (Q', \Sigma, \delta', (0, \emptyset), F')$, where

- $Q' = Q \times 2^Q - F \times 2^{Q-\{0\}}$;
- $F' = \{(i, R) \in Q' \mid R \cap F \neq \emptyset\}$; and
- $\delta'((i, R), a) = (\delta(i, a), R')$, for all $a \in \Sigma$, where

$$R' = \begin{cases} \delta(R, a) \cup \{0\} & \text{if } \delta(i, a) \in F, \\ \delta(R, a) & \text{otherwise.} \end{cases}$$

Then $L(M') = L^2(M)$ and M' has $n2^n - 2^{n-1}$ states.

To show that M' is minimal we will show (a) that all states of M' are reachable, and (b) that the states of M' are pairwise inequivalent with respect to the Myhill–Nerode equivalence relation [7,9].

To prove part (a) let (i, R) be a state of M', where $R = \{r_1, \dots, r_k\}$. If $0 \in R$, assume that $r_k = 0$ and $r_1 < \dots < r_{k-1}$; otherwise, assume that $r_1 < \dots < r_k$. For $j = 1, \dots, k$, define s_j as follows:

$$s_j = \begin{cases} (r_j - 1) \bmod n & \text{if } j = 1, \\ (r_j - r_{j-1}) \bmod n & \text{otherwise.} \end{cases}$$

We first show that if $i = 0$,

$$\delta'\big((0, \emptyset), 1^n(10)^{s_k}1^n(10)^{s_{k-1}} \cdots 1^n(10)^{s_1}\big) = (0, R). \tag{1}$$

For $j = 1, \dots, k$, define R_j as follows:

$$R_j = \left\{1 + \sum_{i=j}^{l} s_i \;\middle|\; j \leqslant l \leqslant k\right\}.$$

Note that $R_1 = R$.

For any state $(0, S) \in Q'$, we have the following facts:

(i) $\delta'((0, S), 1^n) = (0, S \cup \{1\})$,
(ii) if $0 \notin S$, $\delta'((0, S), 10) = (0, S + 1)$, where $+$ is the addition modulo n.

By our choice of the ordering of the r_i's, the R_j's do not contain 0 for $j > 1$. Thus we may apply facts (i) and (ii) inductively to show that for $j = 1, \dots, k$,

$$\delta'\big((0, \emptyset), 1^n(10)^{s_k}1^n(10)^{s_{k-1}} \cdots 1^n(10)^{s_j}\big) = (0, R_j),$$

from which we obtain (1), as required.

If $i > 0$, then let $R' = \{(r_1 - i) \bmod n, \dots, (r_k - i) \bmod n\}$. Just as for $(0, R)$, we see that $(0, R')$ is reachable. Moreover,

$$\delta'\big((0, R'), 1^i\big) = \begin{cases} (i, (R' + i) \cup \{0\}) & \text{if } i = n - 1, \\ (i, R' + i) & \text{otherwise,} \end{cases}$$

which in both cases is equal to $(i, R' + i) = (i, R)$, since by definition $i \in F$ implies $0 \in R$.

To prove part (b) let (i, R) and (j, S) be distinct states of M'. We have two cases.

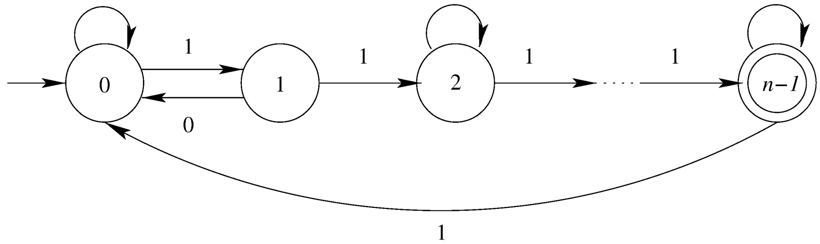

Fig. 1. The DFA M.