Binary Coded Unary Regular Languages

Viliam Geffert

Department of Computer Science, P. J. Šafárik University
Jesenná 5, 04154 Košice, Slovakia
viliam.geffert@upjs.sk

Abstract. $\mathcal{L} \subseteq \{0,1\}^*$ is a *binary coded unary regular language*, if there exists a unary regular language $\mathcal{L}' \subseteq \{a\}^*$ such that a^x is in $\mathcal{L}'$ if and only if the binary representation of x is in $\mathcal{L}$. If a unary language $\mathcal{L}'$ is accepted by an optimal deterministic finite automaton (DFA) $\mathcal{A}'$ with n states, then its binary coded version $\mathcal{L}$ is regular and can be accepted by a DFA $\mathcal{A}$ using at most n states, but at least $1+\lceil \log n \rceil$ states. There are witness languages matching these upper and lower bounds exactly, for each n.
More precisely, if $\mathcal{A}'$ uses $\sigma \geq 0$ states in the initial segment and $\mu \cdot 2^\ell$ states in the loop, where μ is odd and $\ell \geq 0$, then the optimal $\mathcal{A}$ for $\mathcal{L}$ consists of a preamble with at most σ but at least $\max\{1, 1+\lceil \log \sigma \rceil - \ell\}$ states, except for $\sigma = 0$ with no preamble, and a kernel with at most $\mu \cdot 2^\ell$ but at least $\mu+\ell$ states. Also these lower bounds are matched exactly by witness languages, for each σ, μ, ℓ.
The conversion in the opposite way is not always granted: there are binary regular languages the unary versions of which are not even context free.
Removing nondeterminism in a nondeterministic finite automaton (NFA) accepting a binary coded unary language requires at least 2^{n-1} states in the worst case. A conversion of a unary NFA for $\mathcal{L}'$ to an NFA for the binary coded version $\mathcal{L}$ is bounded by $O(n^2)$, and a unary NFA to a binary DFA by Landau's function, with the growth rate $e^{(1+o(1)) \cdot \sqrt{n \cdot \ln n}}$.

Keywords: finite automata · unary regular languages · state complexity