Journal of Automata, Languages and Combinatorics **12** (2007) 1/2, 147–165


# LEARNING UNARY AUTOMATA [1] [2]

GREGOR GRAMLICH [3] and RALF HERRMANN
*Institut für Informatik, Johann Wolfgang Goethe-Universität Frankfurt*
*Robert-Mayer-Straße 11-15, 60054 Frankfurt am Main, Germany*
*e-mail:* gramlich@thi.informatik.uni-frankfurt.de
ralf.herrmann@internet.de

ABSTRACT

We determine the complexity of learning problems for unary regular languages. We begin by investigating the minimum consistent dfa (resp. nfa) problem which is known not to be efficiently approximable within any polynomial, unless $P = NP$. For the case of unary dfa's, we exhibit an efficient algorithm. On the other hand, we show the intractability of the unary minimum consistent nfa problem but provide an efficient quadratic approximation for its optimization version.

The VC dimension for the class of languages accepted by unary dfa's with at most $n$ states is computed as $n + \log n \pm \Theta(\log\log n)$, which (together with the efficient solution for the consistency problem) yields an efficient PAC learning algorithm for this class. We also show that there are no efficient PAC learning algorithms for the class of languages accepted by unary nfa's with at most $n$ states, unless every problem in $NP$ is solvable by a quasipolynomial time Monte-Carlo algorithm. Here we assume that nfa's with few states serve as hypotheses.

In the context of learning with equivalence queries, we consider the number of counterexamples required to learn a unary regular language that is accepted by a dfa with at most $n$ states. When submitting dfa's with at most $n^d$ states ($d \leq n$) as queries, we show the upper bound $O(n^2/d)$ and the lower bound $\Omega((n^2 \cdot \ln d)/(d \cdot (\ln n)^2))$. If only prime cycle lengths $\leq n$ are allowed as queries, we prove that $\Theta(n^2/\ln n)$ counterexamples are necessary and sufficient.

*Keywords:* Unary finite automata, computational learning theory, learning complexity

## 1. Introduction

We investigate the learnability of unary regular languages, i. e., regular languages defined over the alphabet $\Sigma = \{a\}$. In particular we consider *PAC learning* [10], where an unknown concept has to be learned probably approximately correctly from randomly generated positive and negative examples, and learning with *equivalence queries*, where the learning algorithm submits an hypothesis and the teacher either confirms its correctness or replies with a counterexample.

---

[1] Full version of a submission presented at the 7th Workshop on *Descriptional Complexity of Formal Systems* (Como, Italy, June 30 – July 2, 2005).

[2] Partially supported by DFG project SCHN503/2-2.

[3] Corresponding author.