**Master Thesis Proposal**

# Learning Automata over Large Alphabets as an Alternative to Recurrent Neural Networks

**Oded Maler, Nicolas Basset, VERIMAG, Grenbole @univ-grenoble-alpes.fr**

Recurrent neural networks (NN) are to ordinary NN what sequential machines (automata) are to Boolean functions. Using state variables (memory elements) they can encode and learn temporal patterns such as in speech, music and signal processing. In this project we will study an alternative formalism, symbolic automata over large alphabets, to represent and learn classes of sequential behaviors.

In active automaton learning, a target language is given as a black box and by asking membership queries, the learning algorithm constructs an automaton compatible, exactly or approximately, with the language. The well-known $L^*$ algorithm due to Angluin, organizes the sampled words in a table from which the Nerode congruence of the language is constructed

In previous work, we extended this framework to languages defined over large alphabets (bounded subsets of $\mathbb{N}$ or $\mathbb{R}$, Boolean vectors of high dimension), which are represented by symbolic automata, where in each state the outgoing transitions are labeled by predicates that partition the input alphabet into finitely many blocks. The learning procedure has also to learn these partitions using sampling. The goal of the project is to check whether such an approach, in particular a combination of decision tree learning over $\mathbb{B}^n$ and automaton learning, can be an alternative to one or more of the application domains for which recurrent NN are used.

The project will involve the development of new learning algorithms (and improvement of existing ones) and their empirical evaluation, as well as theoretical investigations, such as complexity trade-offs. The relative weights of the theoretical and practical threads will depend on the inclination of the student. One interesting line of research is to consider the case where the state-space of the automaton is also large and represented symbolically and the question is whether one can define (and learn) a symbolic representation of the Nerode congruence.

Further Reading:

O. Maler, I.E. Mens, A Generic Algorithm for Learning Symbolic Automata from Membership Queries, 2017

O. Maler, I.E. Mens, Learning Regular Languages over Large Ordered Alphabets, 2015

D. Angluin, Learning Regular Sets from Queries and Counterexamples, 1987

Do not hesitate to contact us for clarifications.

The work will be conducted at VERIMAG, an internationally recognized lab in verification, embedded systems and hybrid (discrete-continuous, cyber-physical) systems. VERIMAG is located in the nice campus of UGA (Universite Grenobles-Alpes. There will be a possibility to continue for a thesis.