Master Thesis Proposal

Approximate Learning of Monotone Functions

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

A function $f : X \to Y$ between two partially-ordered domains $(X, \preceq)$ and $(Y, \preceq)$ is monotone if $x \preceq x'$ implies $f(x) \preceq f(x')$. We assume a scenario where f is given as a black box that can be queried for values at selected points (active learning) and the goal is to come out with an approximation of h whose distance from f (for example, expected error under uniform distribution over X) is small.

The student will develop such an algorithm (based on initial ideas that we already have) and study the complexity of the problem, both theoretically and empirically. From a theoretical perspective, the complexity of the algorithm will be determined for the general case and for some sub-classes to be identified, for example continuous functions with additional curvature constraints. The algorithms will be implemented and evaluated empirically on different classes of examples.

Partial orders and monotone functions on them constitute very clean and nice objects (especially if you like rectangles), with applications in multi-criteria optimization. Our previous work, on the domain has dealt with monotone partitions, the special case of $f : [0, 1]^n \to \{0, 1\}$, using a multi-dimensional extension of binary search.

Further Reading:

O. Maler, Learning Monotone Partitions of Partially-Ordered Domains, 2017.

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.

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.

Master Thesis Proposal

Temporal Data Mining: Learning from Positive Examples

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

The ultimate goal of the project is to build small models that capture the dynamics of complex and high-dimensional systems, based on sample observations of their behaviors. For example, the system can be an analog circuit C with thousands of transistors and the small abstract model C' will be a low-dimensional differential equation that generates observable behaviors close to those of C. The reduced model can be exported as an approximate description of C and be used for low-cost simulation and verification with other components of the whole system. Another example, coming from our collaboration with Toyota, is a complex car engine which is modeled as a huge network of Matlab/Simulink components. By simulating the system with various input scenarios (driver behavior, road conditions) we obtain a sample S of behaviors (trajectories, signals) for which we would like to build a succinct description, expressed for example as a formula in *signal temporal logic* (STL) or a timed regular expression.

In the context of machine learning, this is a problem of learning from *positive examples*, because we do not observe impossible behaviors. If all we want from C' is to be compatible with S, we have two extreme choices: either let C' be the system that generates/accepts exactly S (known as over-fitting in machine learning) or let C' be the universal system that accepts everything. Both choices are, of course, unsatisfactory and we should choose some other function/set in the sub-lattice between the two.

The project will study this issue first on Boolean functions, where the problem resembles some variant of logic minimization and is concerned with fundamental computer science objects such as terms/cubes, DNF formulas and their sampling. The idea is to develop a learning algorithm that explores the trade-offs between the cardinality of C' and its descriptive complexity (the number of terms in its DNF representation).

After solving the problem for static functions, we will move to functions over Boolean sequences, with the goal of producing a sample-compatible formula in linear-time temporal logic (LTL) or a regular expression. A finite sequence of length k over $\mathbb{B}^n$ can be represented using nk Boolean variables of the form $x_i[t]$. Hence a set/language of sequences can be written as a Boolean function over these variables. One possible research direction would be to characterize temporal functions (sequence classification), those that are expressed compactly by sequential formalisms, as special classes of Boolean functions closed under some variable permutations and not others. Similar investigations can be applied to spatial functions (image classification). The mix of theory and practice (implementation, empirical evaluation) will be determined by the inclination of the student.

Further Reading:

Shai Shalev-Shwartz and Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms,

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.