Learning Automata from their Traces

Reconstructing a dynamic system from its trajectories is an old topic, addressed by several communities. This is called system identification in systems theory, equation discovery in physics, and automata learning in computer science. In CS, one may wish to recover an automaton from words of its language and possibly from counter-examples. Classical "exact" algorithms exist to do so, as the celebrated L-star by D. Angluin, but the latter relies on powerful oracles, i.e. on the possibility to make queries about some words. More recently, new approaches based on artificial neural networks architectures have reached impressive performances on this problem. Beyond performance, they are appealing by several aspects : they are passive methods, relying simply on data-bases of examples (no queries, no need for powerful oracles), they generalize extremely well and can be used as generators, and, focusing on Large Language Models LM, they manage to capture global text features that go beyond classical regularity - spelling, grammar, syntax - as for example style or even meaning. Last but not least, such approximate methods are poorly understood by computer scientists, and proving their reliability remains an issue.

The objective of this internship is to explore the way various NN-based architectures manage to approximate a regular language, produced by some hidden automaton. The focus will be on the way the notion of discrete state emerges in these continuous space systems, and how reliable they are as approximators of automata. Several research directions are envisioned, that will be adapted to the skills and wishes of the candidate. We mention some of them below.

- Exploring the approximation abilities of recurrent neural networks (RNN). RNNs are good approximators of regular languages, but tend to build quasi discrete approximations resembling local automata. This property has to be further understood by examining how well RNN learn more complex languages, and by measuring the distance between the original language and the one approximated by the RNN. This is both an experimental and a theoretical direction, as no algorithms yet exist to estimate such distances.
- Exploring the robustness of the models learnt by RNN, to identify stable regions of their state space and unstable ones. The effect of extra data on such robustness also has to be characterized.
- Replacing a true automaton by its RNN surrogate (used as a generative model for example) raises questions like its reliability. One would like to verify properties of runs produced by such soft automata, for example safety properties. Few algorithms yet exist for model checking such models, and they mostly focus on static feed-forward NN, not recurrent ones.
- Exploring the properties of other architectures. While RNN have a vanishing memory, other structures like GRU or LSTM provide longer term memory. The approximation abilities of such models have to be better understood, in particular to characterize the family of languages they best suit.

The ideal candidate should have a taste for formal methods and abilities for experimental work using standard machine learning libraries.

The internship will take place at INRIA Rennes. The intern will benefit from the standard "gratification" (around 600 euros). Accomodation can be found on the campus.

This research topic can be further extended into a PhD (funding secured).

Bibliography :

- Dana Angluin's L* algorithm : "Learning Regular Sets from Queries and Counter-Examples," 1987, doi :10.1016/0890-5401(87)90052-6
- Frits Vaandrager, Bharat Garhewal, Jurriaan Rot, Thorsten Wissmann : "A New Approach for Active Automata Learning Based on Apartness," 2022, https ://doi.org/10.1007/978-3-030-99524-9_12
- Gail Weiss, Yoav Goldberg, Eran Yahav : "On the Practical Computational Power of Finite Precision RNNs for Language Recognition," 2018, https ://doi.org/10.48550/arXiv.1805.04908
- J. Michalenko, A. Shah, A. Verma, R. Baraniuk, S. Chaudhuri, A. Patel, "Representing Formal Languages : A Comparison Between Finite Automata and Recurrent Neural Networks," ICLR 2019.

Contact : Eric Fabre, eric.fabre@inria.fr, +33 (0)2 99 84 73 26