ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

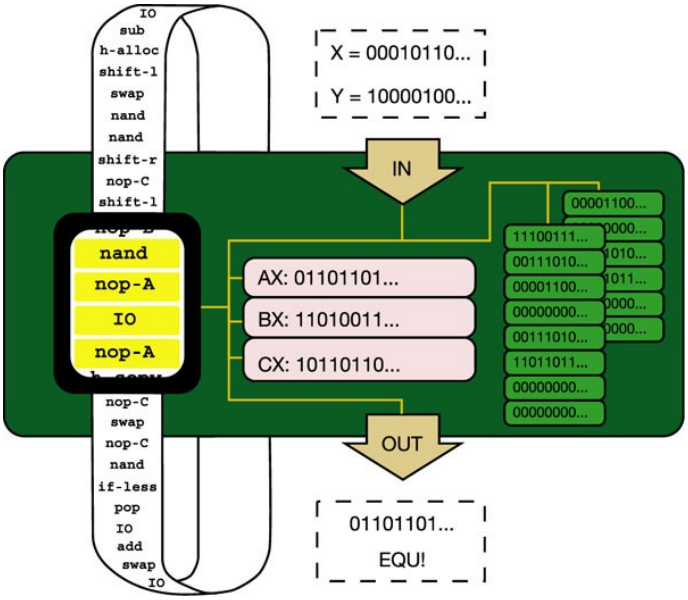
Distributed Computing

Prof. R. Wattenhofer

Benchmark for Sequences

The first number is 2. The second is 4. The third 6. What is the next number? The answer is 14. The solution is the following formula: $x^3 - 6x^2 + 13x - 6$. But if you ask anyone this question, everyone will answer 8. But why 8?

Machine Learning is one of the technologies that has made the most impact in recent years. In particular deep learning has shown some remarkable results.[1] However, deep learning models are data-hungry and seemingly simple problems as above are already difficult for them. This is an old and nascent pursuit in AI. However, with the increase of the computation power this area, similar to deep learning, is becoming more and more approachable.



We have started to build a framework where different methods for searching *simple* algorithms can be implemented. However, gauging the potential of new algorithms requires objectively interpretable, comparable, and reproducible benchmarks. Good benchmarks serve a number of different purposes and groups. For newcomers, benchmarks provide a summary that helps them orient in a maze of new terms, algorithms, and data. For sophisticates, benchmarks provide a baseline. In this thesis, we want to collect data about sequences and label them so that can be used for challenging machine learning tasks. We will start by OEIS database and think how can we create machine learning tasks that are typically hard for the state-of-the-art ML models.

Requirements: Prior experience or a strong interest in programming. Creativity skills are advantageous.

Interested? Please contact us for more details!

Contact

- Ard Kastrati: kard@ethz.ch, ETZ G61.3

[1] Give some examples like, AlphaZero, AlphaFold and so on.