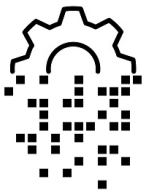
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

Learning Representations for Integer Sequences



Integer sequences are of central importance to the modeling of concepts admitting complete finitary descriptions. In our recent work at DISCO, we introduced a novel view on the learning of such concepts and laid down a set of benchmarking tasks aimed at conceptual understanding by machine learning models. These tasks indirectly assess model ability to abstract, and challenge them to reason both interpolatively and extrapolatively from the knowledge gained by observing representative examples. To further aid research in knowledge representation and reasoning, we presented FACT, the Finitary Abstraction Comprehension Toolkit. The toolkit surrounds a large dataset of integer sequences comprising both organic and synthetic entries, a library for data pre-processing and generation, a set of model performance evaluation tools, and a collection of baseline model implementations, enabling the making of the future advancements with ease.

With FACT in place, exciting new opportunities to leverage the power of deep learning for understanding of finitary abstractions open in front of us.

There are many things that we would like to do. Here are a few, some of which we've already started on.

1. **Large Pre-Trained Model for Abstraction Comprehension.** Essentially a BERT for integer sequences that understands some concepts behind what the sequences represent.
2. **Contrastive Representations of Abstractions.** One way to build representations of FACT data is to apply contrastive learning. We suspect that using this method leads to some loss of knowledge but may also lead to an easier way to identify of prevailing abstractions.
3. **Few-Shot Learning for Abstraction Comprehension.** Can we learn and extrapolate our hypotheses about abstractions from very few examples?
4. **Active Learning for Abstraction Comprehension.** What if our models can actively ask questions about the data we feed them? Could such models be trained to develop better representations of finitary abstractions?

Candidate Profile. Varies from project to project. *Master's* students will work on an extensive project, both coding and advancing our work on the conceptual level. Generally speaking, a good candidate is a competent programmer in the language of his/her choice and is interested in one or more of the following fields: deep learning and deep learning theory, abstraction and reasoning, symbolic computation and symbolic regression.

Interested? Please contact us to learn more!

Contact (please send an email with the following as recipients)

- Peter Belcak: belcak@ethz.ch, ETZ G63