

TAYSIR

TRANSFORMERS+RNN: ALGORITHMS
TO YIELD SIMPLE AND INTERPRETABLE
REPRESENTATIONS

<https://remieyraud.github.io/TAYSIR/>

CALL FOR PARTICIPATION: competition is on!

The Transformers+RNN: Algorithms to Yield Simple and Interpretable Representations (TAYSIR) competition is an on-line challenge on extracting simpler models from already trained neural networks. These neural nets are trained on tasks involving sequences of symbols. Some of these tasks are artificial and some come from real world problems in domains like natural language processing (NLP), bio-informatics, software engineering and others. TAYSIR means "simple" in Arabic.

The quality of the extracted models will be evaluated in two ways:

- How well the extracted model approximates the original model
- The simplicity of the extracted model as measured by assorted metrics

There are two tracks in the competition corresponding to the kind of function the trained neural networks produce.

- Neural nets trained for binary classification. These networks represent functions $\Sigma^* \rightarrow \{0,1\}$. This task can be thought of as extracting models for formal languages.
- Neural nets trained for regression. These networks which represent functions $\Sigma^* \rightarrow \mathbb{R}$. This task can be thought of as density estimation in language modelling or any other situation where sequences of symbols are mapped to real numbers.

Each track consists of about 10 trained models. The trained models are in PyTorch but available in a MLFlow format for compatibility with other frameworks.

The competition has started and will last until April 30th 2023.

Half a day will be dedicated to the competition results during the 16th International Conference on Grammatical Inference to be held in Morocco in July 2023 at the Faculty of Sciences, Mohammed V University in Rabat, Morocco.

<http://www.fsr.ac.ma/icgi2023/>

Participants in TAYSIR will be encouraged to attend ICGI 2023 and to submit an extended abstract presenting their work (2 to 4 pages, including appendices) by May 15th which will be

appended to the proceedings of ICGI (publisher: PMLR) in a track dedicated to the competition. These abstracts will be peer-reviewed primarily for clarity of presentation.

HOW TO PARTICIPATE

Everything can be found on our website: <https://remieyraud.github.io/TAYSIR/>

The submission website for the binary classification Track is here:

<https://codalab.lisn.upsaclay.fr/competitions/11249>

The submission website for the language modeling Track is here:

<https://codalab.lisn.upsaclay.fr/competitions/11683>

You will need to register at the submission websites.

Trained neural nets are provided as MLFlow models. After running your extraction algorithms, participants will upload their extracted model in MLFlow format to the website for evaluation (we provide a submission toolkit to help).

MAIN ORGANIZERS

- Chihiro Shibata, Hosei University, Japan
- Jeffrey Heinz, Stony Brook University, New York, USA
- Rémi Eyraud, Université Jean Monnet, Saint-Etienne, France
- Dakotah Lambert, Université Jean Monnet, Saint-Etienne, France

DEV TEAM

- Aidar Gaffarov, MLDM Mater program
- Badr Tahri Joutei, MLDM Master program
- Mathias Cabanne, Eura NOVA

SCIENTIFIC COMMITTEE

- Ariadna Quattoni, Universitat Politècnica de Catalunya
- Bob Frank, Yale University, USA
- Borja Balle, DeepMind
- François Coste, INRIA Rennes
- Jean-Christophe Janodet, Université Paris-Saclay
- Matthias Gallé, Naver Lab
- Sicco Verwer, TU Delft