

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Martin Vastl
Název práce Deep Learning for Symbolic Regression
Rok odevzdání 2022
Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Roman Neruda **Role** Vedoucí
Pracoviště ÚI AV ČR, v.v.i.

Text posudku:

The thesis deal with a task of symbolic regression, i.e. finding a mathematical formula of an expression given samples of data from this expression. The author utilizes a state-of-the-art neural approach of transformers to solve this problem.

The text of the thesis is well organized, it contains seven parts including introduction and conclusions with future work directions. Chapter one provides basic context and motivation of the problem and approaches used to tackle it. The second chapter is a rather complete overview of related work and previous achievements, dealing with traditional genetic programming approach, reinforcement learning approaches, and the transformer-based approach. In the third chapter, an original method based on transformer networks with local search, is described. The following two chapters deal with experiments of symbolic regression.

The main contributions of this work are:

- The choice of the model itself is very interesting from the methodological point of view. The domain of symbolic regression was dominated by the genetic programming approaches, with new approaches based on neural network architectures appearing recently. The transformer-based approach is novel and worth exploring.
- The original approach of the author provides a complete workflow of the solution dealing not only with the design of transformer network, but also solving related problems by providing local optimization algorithm for constant parameter tuning.
- The experimental part is extensive and very sound. The author reports statistical evaluation of his results, but also deals with other properties such as generalization of the model by means of extrapolation. The hyper-parameter settings and ablation study are very valuable.

I have the following questions for the defense:

- Have you considered other possible approaches such as global optimization to fine-tune the constants of the symbolic regression?
- The initialization of the expressions described in sec. 2.4 have been solved in genetic programming by various procedures for individual generation, such as grow, full or ramped half-and-half heuristics. Have you considered those possibilities?

To conclude, I consider this thesis a high-quality work with original new results deserving a publication in international forum. I gladly recommend the thesis for the defense.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhuje na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum June 1, 2022

Podpis