Co-Design for Training and Verifying Neural Networks

Technische Universität München

Fakultät für Informatik

Lehrstuhl für Echtzeitsysteme und Robotik

Background

The verification of neural networks is essential for their application in systems that have to meet performance guarantees. To obtain all possible outcomes of a neural network for a set of inputs, one can propagate sets through neural networks. Previous works use linear approximations of activation functions, which results in overly conservative results, while nonlinear approximations quickly become computationally infeasible in deep neural networks. We address this issue for the first time by a novel co-design of the training and the verification of neural networks. The main idea is to train the neural network in a way that the verification problem is simplified in the sense that simple approximations of activation functions suffice to verify the neural network.

In order to transfer the verification results to real systems, we additionally add uncertainty to the output of the neural network so that the true outcome is included. It is crucial that the right amount of uncertainty is added so that the results can be transferred without adding to much uncertainty, which would make it impossible to verify the neural network against given specifications.

Description

The main focus of this work is the realization of the novel co-design for training and verifying neural networks. For the conformance checking, existing approaches previously developed at our chair (e.g., [1, 4]) should be integrated in the overall process. Similarly, for the supervised learning of the metal forming process, standard approaches should be implemented, such as those surveyed in [7].

For the co-design of training and verifying the neural network, we envision to not only minimize the average squared error between the output of the network and the desired output, but also the size of the output set when adding uncertainty to the input of the neural network. Computing an over-approximation of the output set would be computationally expensive and is not required to robustify the neural network – only the fully trained neural network has to be formally verified. For this final verification, we will use our previous work on over-approximative set propagation of neural networks as a basis [2, 3]. Instead, we will use efficient approximations for the robustification of neural network using a) sensitivity analysis, b) reachability analysis without approximation errors (see [6]), and c) approximated approximation errors (see [5]).

Tasks

- Implementation of supervised learning to model metal forming processes.
- Formalization of specifications for a specific metal forming process with a Bihler machine at the chair of Metal Forming and Casting.
- Integration of the approximate size of output sets in supervised learning.
- Reachset conformance checking of the resulting neural network using set propagation and additive uncertainties.
- Formal verification of the resulting neural network using set propagation.
- Optional: order reduction of the used set representation.
- Optional: consideration of multiplicative uncertaintes for conformance checking rather than additive uncertainties.
- Implementation of the developed approaches in CORA.

Supervisor:
Prof. Dr.-Ing. Matthias Althoff

Advisor:
Prof. Dr.-Ing. Matthias Althoff, M.Sc. Tobias Ladner

Research project:
-

Type:
MA

Research area:
Formal verification, conformance checking, supervised learning

Programming language:
MATLAB

Required skills:
Knowledge in formal methods and machine learning, good mathematical background

Language:
englisch

Date of submission:
23. März 2023

For more information please contact us:

Phone: +49.89.289.18134

E-Mail: althoff@tum.de

Internet: www6.in.tum.de

References

[1] N. Kochdumper, A. Tarraf, M. Rechmal, M. Olbrich, L. Hedrich, and M. Althoff. Establishing reachset conformance for the formal analysis of analog circuits. In *Proc. of the 25th Asia and South Pacific Design Automation Conference*, page 199–204, 2020.

[2] Niklas Kochdumper, Christian Schilling, Matthias Althoff, and Stanley Bak. Open- and closed-loop neural network verification using polynomial zonotopes. In *Proc. of NASA Formal Methods*, 2023.

[3] Tobias Ladner and Matthias Althoff. Automatic abstraction refinement in neural network verification using sensitivity analysis. In *Proc. of the 26th ACM International Conference on Hybrid Systems: Computation and Control*, 2023.

[4] S. B. Liu and M. Althoff. Reachset conformance of forward dynamic models for the formal analysis of robots. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, page 370–376, 2018.

[5] Lukas Schäfer, Felix Gruber, and Matthias Althoff. Scalable computation of robust control invariant sets of nonlinear systems.

[6] B. Schürmann and M. Althoff. Optimizing sets of solutions for controlling constrained nonlinear systems. *IEEE Transactions on Automatic Control*, 66(3):981–994, 2021.

[7] D. Weichert, P. Link, A. Stoll, S. Rüping, S. Ihlenfeldt, and S. Wrobel. A review of machine learning for the optimization of production processes. 104:1889–1902, 2019.

Technische Universität München

Fakultät für Informatik

Lehrstuhl für Echtzeitsysteme und Robotik