Modular Verification of Neural Networks

Topic for "Praxis der Forschung"
Philipp Kern, Samuel Teuber

Background I: Neural Network Verification

- **Question:** Does NN \( g \) respect specification \( \psi (x, y) \)?
- **Problem:** Verification is NP-complete even for piece-wise linear feed-forward NNs
- Current approaches do not scale to State-of-the-Art networks

Background II: Traditional Program Verification

- Break down large program into smaller components
- Verify properties on components
- Compose properties to prove correctness of complete program

Idea: Modularize the verification process for NNs

Infer properties \( \phi_1, \phi_2 \) such that:

- \( \phi_1 (x, h^{(2)}) \) holds on \( x \rightarrow h^{(2)} \)
- \( \phi_2 (h^{(2)}, y) \) holds on \( h^{(2)} \rightarrow y \)
- \( \phi_1 (x, h^{(2)}) \land \phi_2 (h^{(2)}, y) \implies \psi (x, y) \)

In particular: Application to recursive neural networks (LSTMs)

Relevant Literature:

- Data-Driven Invariants [2, 4, 1]
- NN Invariants [5, 6, 3]
- Modular Bounded Model Checking [7]

---


