Verification of Neural Networks Using Symbolic Interval Propagation

Shirley Xu, shirley.xu.23@bishops.com
The Bishop's School, Class of 2023
USC Viterbi Department of Electrical Engineering, SHINE 2021

Introduction and Impact

With advancements in machine learning, the algorithms behind self-driving vehicles are becoming more prominent and more advanced. However, research has shown that there are still many corner cases, such as minor changes in the amount of noise or lighting, that lead to fatal errors by neural networks *[1].* For example, there have been numerous fatal reports of Teslas crashing into white-colored trucks (the algorithm mistaking the truck for clouds in the sky) while on autopilot.

Our research aims to verify these systems using Symbolic Interval Propagation. By effectively approximating the behavior of the ReLU, symbolic interval propagation helps determine the type of a neuron faster, which accelerates the verification.

The flow chart of the verification is shown in Figure 1. If the safety property is satisfied, then the model is safe. If not, a counterexample must be found to prove the model is not safe.

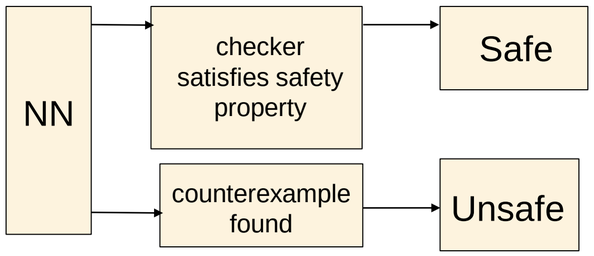

Figure 1: Algorithm flow of the verification of a neural network. PC: Shirley Xu

Introduction (cont.)

In the six context examples below with increasing noise, the model detects a "9". Although it is not difficult for a human to determine that the right answer, the network has a lot of trouble ignoring the noise.

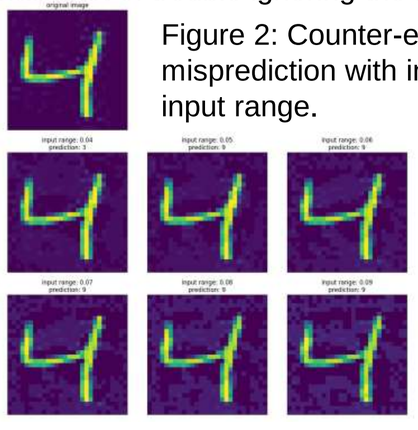

Figure 2: Counter-example of misprediction with increasing input range.

PC: Kevin Chiaming Chang

Symbolic Interval Propagation

- A technique that computes output interval bounds of a neural network
- Restrains nonlinear part of neuron's output → z = ReLu(Eq) with two linear equations: Upper equation and lower equation of a neuron achieves tightest possible bound due to its least maximum distance from Eq

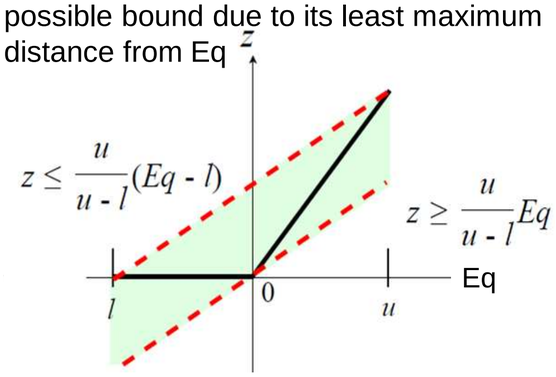

Figure 3: Symbolic Linear Relaxation of Non-Linear ReLu Function. [1]

Experimental Results

- Neuron type decided by input range
- Active: Lower Bound > 0
- Indeterminate: Upper Bound > 0 > Lower Bound
- Inactive: Upper Bound < 0
- As the input range of a neuron expands, the neuron is more likely to be an indeterminate neuron

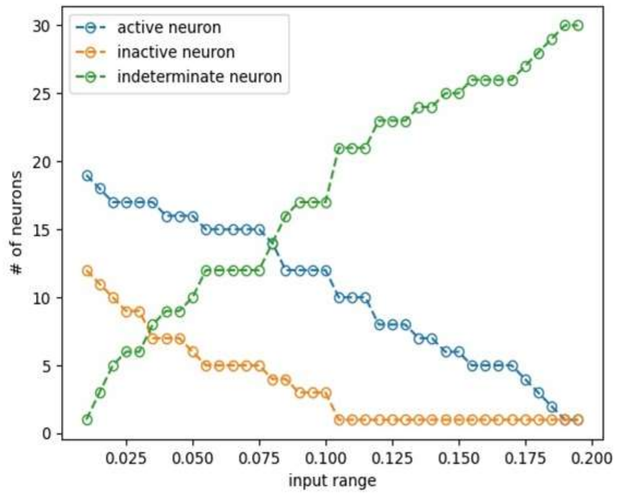

Figure 4: Different types of neurons and input range. PC: Shirley Xu

- Number of active and inactive neurons: inverse relationship with input range
- Indeterminate neurons: direct relationship with input range.

Relevance to STEM Coursework

Next year, I will be taking Advanced Honors Computer Science at my school, which focuses on machine learning and research. I believe my experience at SHINE has given me a deeper understanding on neural networks and machine learning techniques in the field.

Advice for Future SHINE Students

SHINE is a rare experience to have. Enjoy every meeting you have, whether it be all-cohort meetings or meetings with your mentor. Always be flexible and open to trying new things or methods you may not have thought of the first time around. Understand that it is okay to be wrong. Your professor, mentor, center mentor, and everyone else at SHINE are there to help you learn and further your experience.

Skills Learned

- Ubuntu VirtualBox- Virtual Environment on Windows
- Applications of NumPy Arrays
- Deep learning neural network

Citations

[1]: *Pei, Kexin, et al. "Efficient Formal Safety Analysis of Neural Networks." ArXiv.org, 7 Nov. 2018, arxiv.org/abs/1809.08098.*

Acknowledgements

To Kevin, my mentor, Professor Nuzzo, Ms. Emily, my center mentor, and Dr. Mills, a big thank you for helping make my SHINE experience amazing. I've learned a lot and it would not be possible without you all. Thank you so much!