Research progress of neural network repairing

Reporter: Chi Zhiming

May 20, 2023

目录

1. Existing research

2. Preliminary Study

3. Future Work

目录

1. Existing research
2. Preliminary Study
3. Future Work

Art

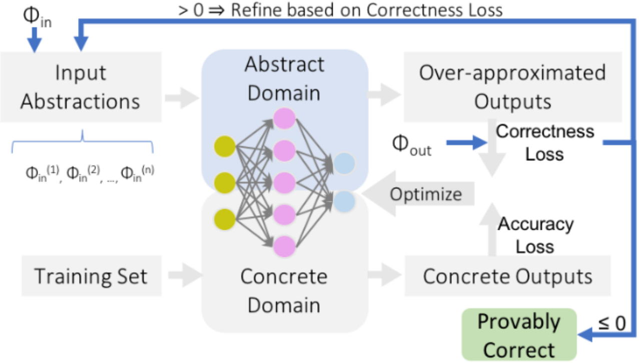

Fig. 1: The ART framework.

REASSURE

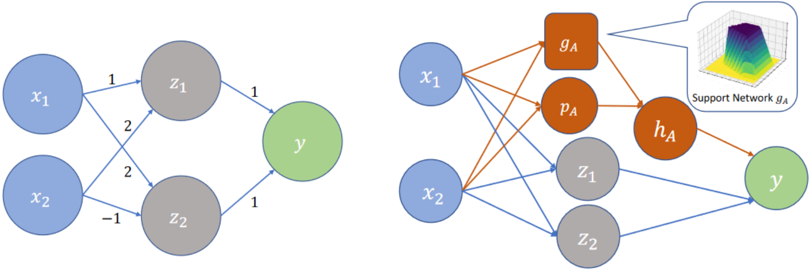

Figure 2: Left: the target DNN with buggy inputs. Right: the REASSURE-repaired DNN with the patch network shown in red. Support network $g_{\mathcal{A}}$ is for approximating the characteristic function on $\mathcal{A}$; Affine patch function $p_{\mathcal{A}}$ ensures the satisfaction of Φ on $\mathcal{A}$; The design of the patch network $h_{\mathcal{A}}$ ensures locality for the final patch.

目录

1. Existing research
2. Preliminary Study
3. Future Work

The challenging of REASSURE

- The number of activation patterns is exponential.
- It can not repair the mutiple properties.

Our idea

- Repairing regions: polytopes → regions partitioned by NN.
- Patch networks: linear function → NN.
- Framework: leverage the framework of Art to training the support network and patch network until the NN satisfy the desired properties.

Current progress

- Be familiar with the code of ART and REASSURE
- Encoding the support networks and patch networks to the framework of ART.

目录

1. Existing research
2. Preliminary Study
3. Future Work

Future Work

- Construct the one-to-one correspondence between support networks and patch networks
- The number and framework of support networks and patch networks: hard-coding → heuristic
- Research the performence when we train the repaired network with fixing the original parameters of NN rather then not fix them.
- More properties, such as fairness and robustness.

Thank you