A programming model of verification of neural network

迟智名

September 27, 2022

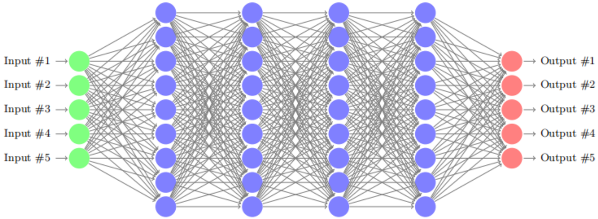
目录

目录

Reference paper

- Integrating Simplex with DPLL(T)
- Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks

Neural network model

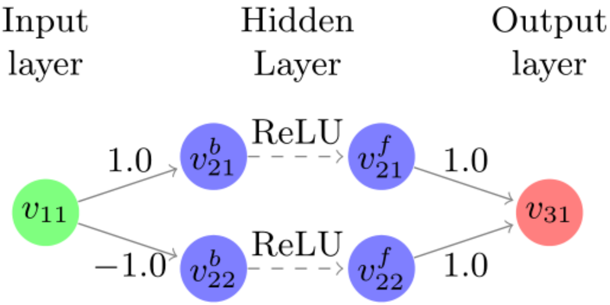


- s_i: the size of layer i.
- $v_{i,j}$: The value of the j-th node of layer i.
- V_i: the column vector $[v_{i,1}, ..., v_{i,si}]$.
- Relu(x): max(0,x)
- $V_i = ReLU(W_i V_{i-1} + B_i), W_i : s_i \times s_{i-1}, B_i : s_i$

目录

verification of neural network

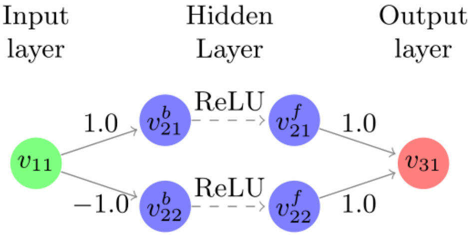
A small example:



To verify Property:

- $v_{11} \in [0, 1]$
- $v_{31} \in [0.5, 1]$

Programming model



To verify Property:

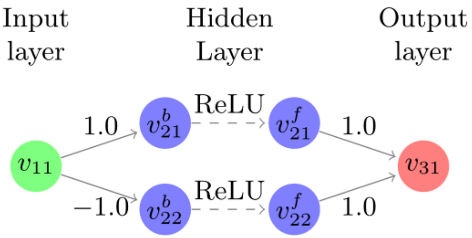
- Verify: $v_{11} \in [0, 1]$ and $v_{31} \in [0.5, 1]$
- equation constraints:

$$\begin{cases} a_1 = -v_{11} + v^b_{21} \\ a_2 = v_{11} + v^b_{22} \\ a_3 = -v^f_{21} - v^f_{22} + v_{31} \end{cases} \tag{1}$$

- bound constraints:

$$\begin{cases} a_{1,2,3} = 0 \\ v^b_{21}, v^b_{22} \in (-\infty, \infty) \\ v^f_{21}, v^f_{22} \in (0, \infty) \end{cases} \tag{2}$$

Programming model



Split constraints:

- $(v_{21}^b \geq 0 \implies v_{21}^f = v_{21}^b) \wedge (v_{21}^b \leq 0 \implies v_{21}^f = 0)$
- $(v_{22}^b \geq 0 \implies v_{22}^f = v_{22}^b) \wedge (v_{22}^b \leq 0 \implies v_{22}^f = 0)$

Programming model

- equation constraints:

$$\begin{cases} a_1 = -v_{11} + v_{21}^b \\ a_2 = v_{11} + v_{22}^b \\ a_3 = -v_{21}^f - v_{22}^f + v_{31} \end{cases} \tag{3}$$

- bound constraints:

$$\begin{cases} v_{11} \in [0,1], v_{31} \in [0.5,1] \\ a_{1,2,3} = 0 \\ v_{21}^b, v_{22}^b \in (-\infty, \infty) \\ v_{21}^f, v_{22}^f \in (0, \infty) \end{cases} \tag{4}$$

- Split constraints:

$$\begin{cases} (v_{21}^b \geq 0 \implies v_{21}^f = v_{21}^b) \wedge (v_{21}^b \leq 0 \implies v_{21}^f = 0) \\ (v_{22}^b \geq 0 \implies v_{22}^f = v_{22}^b) \wedge (v_{22}^b \leq 0 \implies v_{22}^f = 0) \end{cases} \tag{5}$$

Discussion: NPC-problem

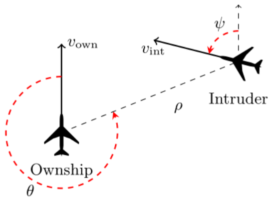
- Note the split constraints:

$$\begin{cases} (v_{21}^b \geq 0 \implies v_{21}^f = v_{21}^b) \wedge (v_{21}^b \leq 0 \implies v_{21}^f = 0) \\ (v_{22}^b \geq 0 \implies v_{22}^f = v_{22}^b) \wedge (v_{22}^b \leq 0 \implies v_{22}^f = 0) \end{cases} \tag{6}$$

- n nodes $\rightarrow 2^n$ conditions
- this theoretical worst-case behavior is also seen in practice

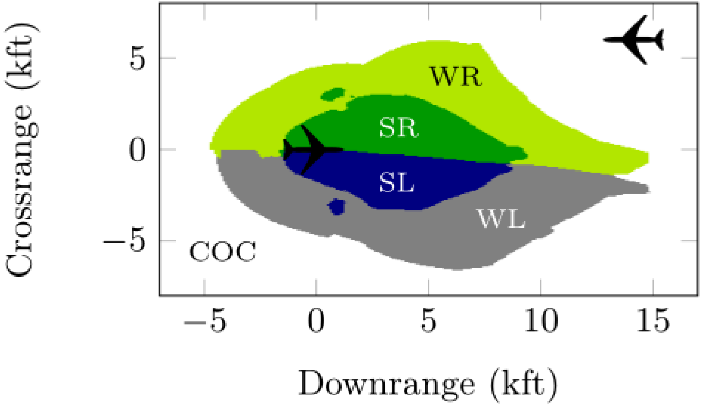
目录

ACAS Xu(Airborne Collision Avoidance System X unmanned)



- Input of networks: $v_{own}, v_{int}, p, \theta, \phi$
- 45 networks: 5 previous advisories multiplies 9 discretizing(Time until loss of vertical separation)
- 5 outputs: SL,SR,WL,WR,COC(Clear-of-Conflict)

Prove Network Properties



Thank you