INF575 Reading Assignment: Verisig & Verisig 2.0

Verifying Neural Networks as Hybrid Systems

Yee-Jian Tan

École Polytechnique

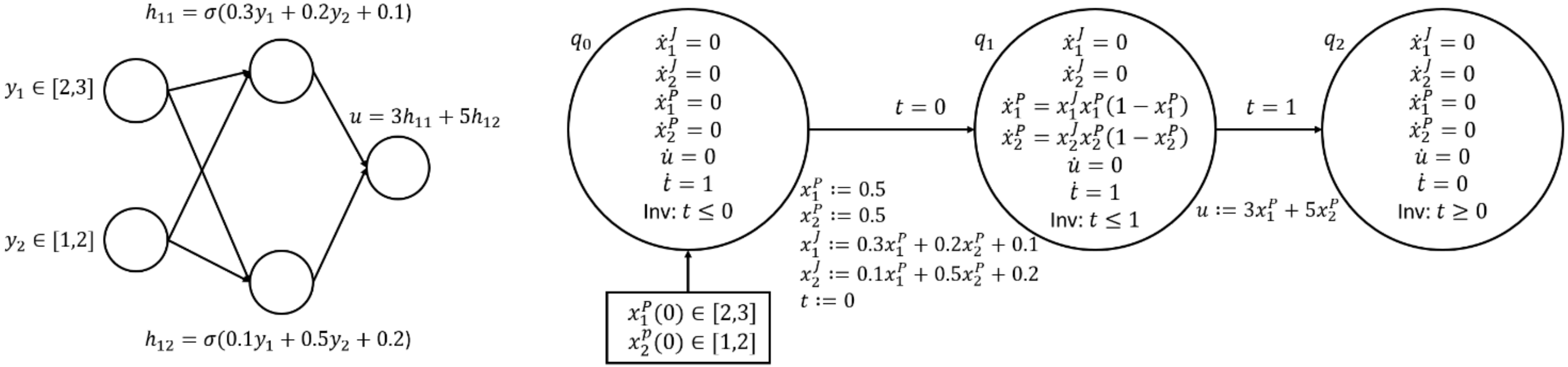
December 19, 2023

Plan for today

- What is Verisig? 3
    - Why can we do that? 4
    - Why should we do that? 7
    - How can we do better? 9
    - Possible Limitations 10

What is Verisig?

- Transforms a Neural Network into an equivalent Hybrid System.



(a) Example DNN. **(b) Equivalent hybrid system.**

Figure 2: Small example illustrating the transformation from a DNN to a hybrid system.

Why can we do that?

- Sigmoid functions are solutions to quadratic differential equations.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Why can we do that?

- Sigmoid functions are solutions to quadratic differential equations.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\frac{d\sigma}{dx}(x) = \sigma(x)(1 - \sigma(x))$$

Why can we do that?

- Sigmoid functions are solutions to quadratic differential equations.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\frac{d\sigma}{dx}(x) = \sigma(x)(1 - \sigma(x))$$

$$\frac{\delta g}{\delta t}(t, x) = \dot{g}(t, x) = xg(t, x)(1 - g(t, x)).$$

- Then treat a neuron as a hybrid system, and analyze using Taylor Models.

Why should we do that?

- Verification of property is decidable for one layer

  Reason: it is a $\mathcal{R}$-formula: $(\mathbb{R}, <, +, -, \cdot, 0, 1)$

Why should we do that?

- Verification of property is decidable for one layer

  Reason: it is a $\mathcal{R}$-formula: $(\mathbb{R}, <, +, -, \cdot, 0, 1)$

- δ-decidable for multiple layers

  Reason: it is a $\mathcal{R}_{\exp}$-formula: $(\mathbb{R}, <, +, -, \cdot, 0, 1, \exp)$ since we don't know how to eliminate the e^{-x}.

How can we do better?

1. Taylor Model Preconditioning

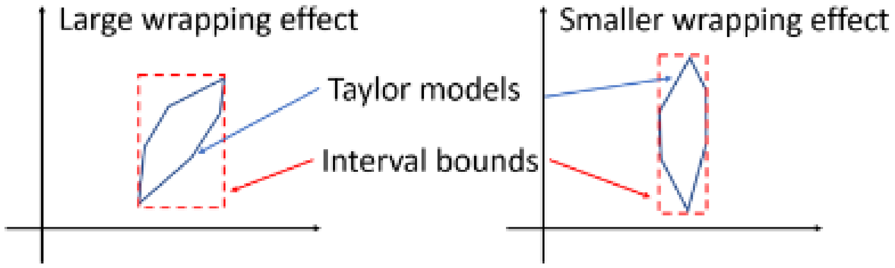


Fig. 2. The wrapping effect for different taylor model orientations.

2. Shrink Wrapping

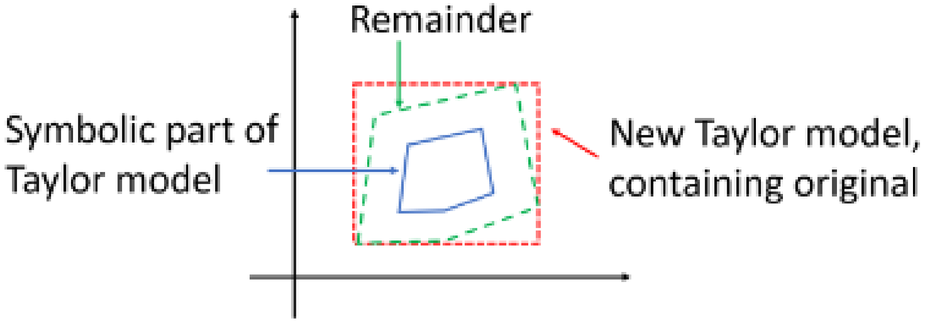

Remainder
Symbolic part of Taylor model
New Taylor model, containing original

Fig. 3. Illustration of the shrink wrapping method.

3. Parallelism: one neuron one core.

Possible Limitations

- Elimination of remainder: reduces computation overhead, but increases inaccuracies

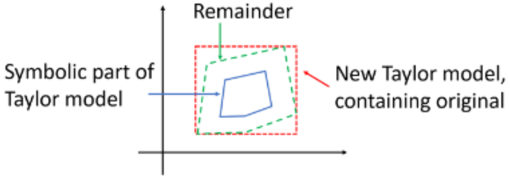

Remainder
Symbolic part of Taylor model
New Taylor model, containing original

Fig. 3. Illustration of the shrink wrapping method.

- Experiments in Verisig 2.0 have very few layers (2-3), which is where the sampling method could shine due to less overhead.