Formal Verification of Neural Network Behaviour for Stability Assessment in Modern Distribution Grids

Master's Thesis Proposal

Haodong Luo

Supervisor: Dr. Atanasious

June 26, 2025

Motivation & Problem Statement

The Potential and Pitfalls of Neural Networks

- **Potential:** Significant potential for accelerating complex tasks like power system security assessment.
- **Pitfall:** Their "black box" nature is a major barrier to adoption in safety-critical applications. Even high-accuracy networks can be vulnerable to small **adversarial examples** causing misclassification.

A New Challenge: The Modern Distribution Grid

- High penetration of volatile Photovoltaics (PV) and uncertain Electric Vehicle (EV) charging loads introduces new **voltage stability** and **congestion management** challenges.
- Using NNs for rapid assessment is a promising approach, but their **reliability in this new context is unknown**.

Goal of This Research

To apply a state-of-the-art formal verification framework to this new, critical problem domain, providing **provable guarantees** for NN reliability.

Novelty & Contribution: Exploring a New Frontier

Comparison Dimension	Foundation Paper (Venzke et al., 2020)	My Proposed Thesis
System	Transmission Grid	**Distribution Grid**
"Safety" Criteria	Static/Dynamic Security (N-1)	**Time-Series Operational Feasibility** (24h voltage/thermal limits)
NN Input ('x')	Operator Control Variables (Generator Dispatch)	**Planning/Uncertainty Parameters** (PV/EV Penetration)
Research Goal	Verify tools for real-time **operational decisions**	Verify tools for **planning analysis** under uncertainty

Core Contribution

To adapt and apply a proven formal verification framework to a new, more complex, and highly relevant problem domain.

Key Research Questions

1. **Quantifying Robustness:**
   For an NN trained to classify the 24-hour operational feasibility of a distribution grid, what is its **adversarial accuracy** against perturbations in PV/EV penetration forecasts?

Key Research Questions

1. **Quantifying Robustness:**
   For an NN trained to classify the 24-hour operational feasibility of a distribution grid, what is its **adversarial accuracy** against perturbations in PV/EV penetration forecasts?
2. **Improving Verification Efficiency:**
   How does using **weight sparsification** on the NN affect the computational time of the MILP-based verification, and what is the trade-off with classification accuracy?

Key Research Questions

1. **Quantifying Robustness:**
   For an NN trained to classify the 24-hour operational feasibility of a distribution grid, what is its **adversarial accuracy** against perturbations in PV/EV penetration forecasts?
2. **Improving Verification Efficiency:**
   How does using **weight sparsification** on the NN affect the computational time of the MILP-based verification, and what is the trade-off with classification accuracy?
3. **Enhancing NN Reliability:**
   Can **adversarial retraining**—retraining the NN on systematically identified adversarial examples—measurably improve its robustness and predictive performance on unseen data?

Proposed Methodology

1. Data Generation & NN Training

- **Tool:** OpenDSS or Pandapower
- **Case:** IEEE 13-bus system
- **Dataset:** Latin Hypercube Sampling, 24h time-series labeling
- **Training:** TensorFlow/PyTorch

2. Formal Verification & Analysis

- **Core:** Reformulate NN as an **MILP**
- **Robustness Check:** Solve MILP to find adversarial perturbations
- **Enhancement:** Test weight sparsification & adversarial retraining

Data Generation → NN Training → MILP Verification → Result Analysis

Expected Outcomes & Key Figures

- **Figure 1: Adversarial Accuracy Plot**
  (The core evidence quantifying NN vulnerability)
- **Figure 2: Security Boundary Visualization**
  (Visually demonstrating the flaws in the NN's learned boundary)
- **Table 1: Performance Comparison: Dense vs. Sparse NN**
  (Quantifying the verification speed-up from sparsification)
- **Figure 3: Robustness Improvement from Retraining**
  (Proving the effectiveness of the enhancement method)

Scope of Work & Timeline

Scope of Work

- **Must-Have (Core Thesis):** Implement the full data generation and baseline verification pipeline for one test system (e.g., IEEE 13-bus
- **Should-Have (Strong Contribution):** Implement and benchmark a sparse NN for verification efficiency and perform adversarial retraining to show robustness improvement.
- **Could-Have (Ambitious Extensions):** Try different ways of MILP setting.