ROBUSTNESS VERIFIER HEURITICS FOR NEURAL NETWORKS

R. Deliallisi

ETH Zürich, D-INFK
Rämistrasse 101, 8092 Zurich, SWITZERLAND

*C. Trassoudaine**

IMT Atlantique,
655 Avenue du Technopôle, 29280 Plouzané, FRANCE
EURECOM, Data Science dpt.,
450 route des Chappes, 06410 Biot, FRANCE

Introduction

The purpose of this work is to prove formally the local-robustness of neural-networks (NN) by heuristically combining Box analysis and linear-programming (LP) solving. In this paper, we aim to present a time-efficient way of verifying NN robustness large perturbations $\eta = \{\epsilon_0, ..., \epsilon_n\}$. We also make the simplification that the perturbation range is the same for each input neuron $n_{-1,i}$ such that the perturbed input $\hat{n}_{-1,i} = n_{-1,i} \pm \epsilon, \forall i \in \{1, ..., n\}$.

1. ANALYSIS TECHNIQUES

1.1. Box analysis

A very simple and fast approach to solve the NN robustness problem is to use a polyhedra abstract domain as defined in [1]

1.2. Linear programming

1.2.1. Range analysis

1.2.2. Robustness verification

2. HEURISTICS

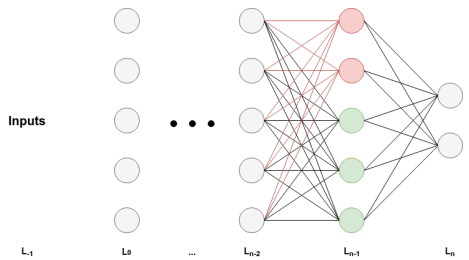

Fig. 1. Last layers additionnal optimization

3. RESULTS

*Work performed while at ETH Zürich
[1] **AI2**.