

VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wānanga o te Ūpoko o te Ika a Māui



School of Engineering and Computer Science
Te Kura Mātai Pūkaha, Pūrorohiko

PO Box 600
Wellington
New Zealand

Tel: +64 4 463 5341
Fax: +64 4 463 5045
Internet: office@ecs.vuw.ac.nz

Logical Neural Networks: Opening the black box

Daniel Thomas Braithwaite

Supervisor: Marcus Frean

Submitted in partial fulfilment of the requirements for
Bachelor of Science with Honours in Computer Science.

Abstract

Neural networks have been shown to be universal function approximators and have been employed to solve a number of non-trivial problems. The downside to these networks is that they are very difficult (if not impossible) to understand. We aim to show that building NN's out of Noisy logic gates are also universal function approximators, have comparable performance to regular NN's and the added benefit of being understandable to humans

1. Introduction

Neural Networks perform exceptionally well over a wide range of different problems. However once trained these networks become a black box, near impossible for a human to understand what features the network is using to solve the problem presented to it. Logical Neural Networks (neural networks with logical activation functions) have been shown to provide a more understandable representation, however LLN's have been neglected not a lot of effort has been put into developing them.

2. The Problem

Research done previously into LLN's has shown that using a network with layers consisting of Noisy-OR and Noisy-And neurons can achieve a reasonable accuracy on the MNIST dataset while also providing a more understandable network. To begin this project we take a step back to ensure we have the right approach. A network built out of neurons that resemble logic gates we would expect would be able to learn any given boolean formula, if we can't achieve this then there is a fundamental problem with our approach to LLN's.

Ideally we would give a mathematical proof showing that LLN's are (or are not) universal boolean function approximators. If LLN's are this would provide confidence for them also being universal function approximators in the general case. If we aren't able to prove (or provide evidence via experiments) that LLN's are universal function approximators then their uses would be very limited.

3. Proposed Solution

Before we can start we must address a fundamental issue with our existing setup. Namely our choice of "gates" is not a functionally complete set. Trying to learn a boolean function with only AND and OR might work but it will severely limit us. We propose to implement a number of different sets of functionally complete neurons, such as NAND, NOR. Ideally we would find that they all give the same performance but given we are modeling our discrete gates with a continuous function we could see unexpected results.

After experimenting with LLN's for approximating boolean functions we would have a good idea for whether they are universal boolean function approximators or not, but intuitively it would make sense for them to be, given that we would be using a functionally complete set of logic gates (which we know in their discrete form can represent any boolean expression). However as discussed before, given we are approximating our discrete gates as continuous the same principles may not hold, for this reason we would first need to investigate what effects our continuous model has on the properties of logic gates.

4. Evaluating your Solution

5. Resource Requirements

