Summary of the NIPS '90 workshop on comparing Decision Trees and Neural Networks

Lorien Y. Pratt*

This document describes a workshop on comparing decision trees and neural networks which was held in conjunction with NIPS '90 on November 30, 1990 in Keystone, Colorado. The workshop handout is also available, by anonymous ftp from cheops.cis.ohio-state.edu; its name is pratt-dtnnet-workshop-handout.ps. That document contains references to relevant papers, as well as a summary of comparative points between neural networks and decision trees.

Since 1989, at least six papers have been published which discuss empirical and theoretical comparisons between neural network and decision tree learning algorithms. In addition, there have been a number of studies which have explored issues at the interface between these two research areas. The goal of this workshop was to bring together researchers from both areas in order to address some important questions about their relationship. A number of invited speakers gave short talks summarizing their work.

First, in my introductory remarks, I presented a historical overview of the relationship between these two fields, and pointed out that comparisons between them fall on the disjoint axes of representation, learning algorithm, and performance algorithm.

Next, Tom Dietterich, Les Atlas, and Ah Chung Tsoi described several empirical studies. Although one general conclusion of our workshop was that there is no definitive answer to whether decision trees or networks are "best" in general, these speakers did give a body of evidence which showed that, for certain tasks, back-propgation networks out-performed decision tree induction algorithms like CART, C4.5, and ID3. The tasks explored included large and difficult real-world sized problems such as power system security. Thus, it seems to be in the interest of persons applying decision tree technology to consider a neural network approach, based on the potential improvement in performance.

Next, Paul Utgoff and Ananth Sankar each presented their work on architectures which combine decision trees with feedforward neural networks. In their schemes, a decision tree is constructed until a class distinction seems solvable by a network, at which point a small linear network is placed at the decision tree leaf. There is reason to believe that this model has considerable computational advantages, and both Sankar and Utgoff presented preliminary empirical results which are suggestive that this kind of hybrid technique may give superior performance in some situations.

Stephen Omohundro presented a unifying view of neural networks and decision trees by placing them both within a paradigm of model selection in a Bayesian framework. Terry Sanger instantiated this idea into a general method using a mechanism called a "basis function tree"; CART, MARS, and decision trees are special cases of this general model.

*pratt@paul.rutgers.edu, Rutgers University Computer Science Department, New Brunswick, NJ 08903. This work was performed in part while the author was supported by a research assistantship from Siemens, Inc.

In a discussion period, we debated several conclusions which might be drawn on the relationship between decision trees and neural networks. These are outlined below:

- There is no single "best" method in all situations. The superiority of one or the other of neural networks or decision trees is problem- (or perhaps problem-class)-dependent.
- Today, after there have only been a couple of years in which decision trees and neural networks have been directly compared, it is premature to expect a clear picture of when one is superior over the other. In other research areas, expert practitioners have a number of techniques at their disposal, and some ability to judge *a priori* which is better for a particular task. We are moving towards the ability to have such expertise about classification and regression tasks, but the picture is still fuzzy when it comes to knowing whether a decision tree or network approach is superior.
- Complementary to the idea of developing a body of techniques for solving classification and regression tasks, it might be helpful to develop a *problem*-oriented taxonomy, focusing on the aspects of problems that make them amenable to one or the other approach.
- We were generally in agreement with a "strong convergence hypothesis" – that these fields are close enough to each other that it is important for neural network practitioners to have expertise in decision trees, and vice versa.
- Not only has there been intensional work to build research projects which bridge the gap between neural networks and decision trees, but some of the primary thrusts in research today can be characterized as the movement towards some sort of compromise between these fields. Neural networks were originally formulated as having strictly global nodes, and the ability to form distributed representations with them was emphasized. Decision trees were formulated with local nodes, and their ability to be understood by humans was of primary importance. Today, both fields see not only the benefits but also some potential disadvantages of these extreme stances, and so we are seeing decision trees with global nodes, and lots of recent interest in networks with local nodes, as well as those which have nodes to which a symbolic meaning can be attributed.
- Paul Utgoff pointed out that it's important to control for any user intervention (via parameters like learning rate and momentum) when measuring the performance of trained networks, especially when their performance and/or learning times are being compared to decision trees, which have fewer user-modifiable parameters.

  The general issue of measurement methodology was raised several times at our workshop; the recent book by Weiss and Kulikowski[1] addresses these issues in detail, as well as giving an excellent introduction to decision trees and neural networks.

We'd like to thank the speakers and the 50-odd workshop participants who, by bringing their experience from diverse applied and research areas, served to keep the discussion lively. We believe that questions of the relationship between decision trees and neural networks will continue to enjoy active debate for a few years to come, and that this workshop is just one landmark in a complex, but fascinating, territory.

[1] "Computer Systems that Learn", Morgan Kauffman, 1990