



Published in Towards Data Science



Barak Or, PhD

[Follow](#)Oct 29, 2022 · 4 min read · [Listen](#)

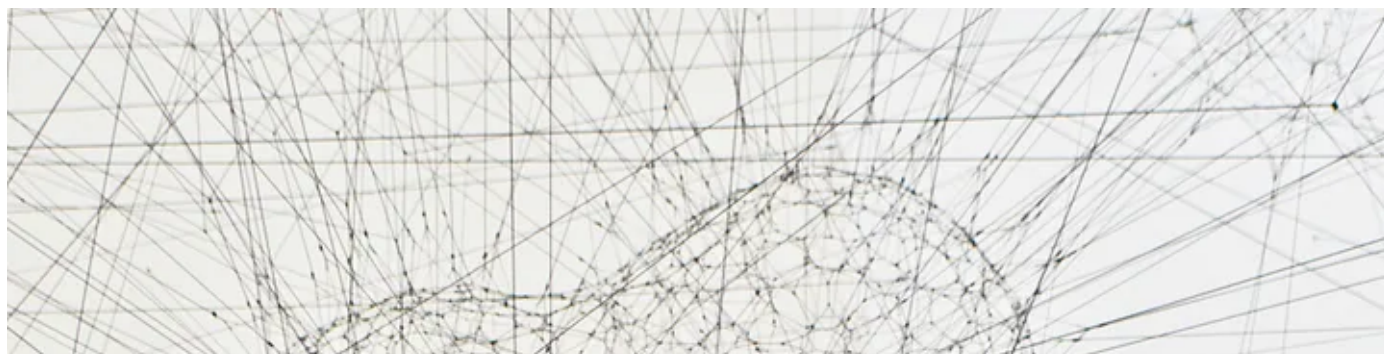
Save



Pushing Explainable AI: Neural Networks Are Decision Trees

Exploring a new paper that aims to explain DNN behaviors

Recently, a great researcher from AAC Technologies, Caglar Aytekin, published a paper titled “Neural Networks are Decision Trees.” I read it carefully and tried to understand exactly what the big discovery from this paper is. As many data scientists will probably agree, many transformations take one algorithm into another. However, (deep) neural networks (DNNs) are hard to interpret. So, did Aytekin discover something new that leads us one step closer to the explainable AI era?

[Open in app](#) ↗[Sign up](#)[Sign In](#)

Search Medium



unsplash

In this post, let's explore the paper and try to understand if this is actually a new discovery. Alternatively, we'll examine if it is just an important spotlight that any data scientist needs to know and remember while handling the DNN interpretability challenge.

Introduction and Background

Aytekin demonstrated that any classical feedforward DNN with piece-wise linear activation functions (like ReLU) can be represented by a decision tree model. Let's review the main difference between the two:

*DNN fits parameters to **transform** the input and indirectly direct the activations of their neurons.*

*Decision trees **explicitly fit** parameters to direct the data flow.*

The motivation for this paper is to **tackle the black-box nature of DNN models** and have another way to explain DNN behaviors. The work handles fully connected and convolutional networks and presents a directly equivalent decision tree representation. So, in essence, it examines the transformation from DNN to a decision tree model when taking a sequence of weights with non-linearity between them and transforming it into a new weights structure. One additional result that Aytekin discusses is the advantages of the **cr** **96** **3** **N** in terms of computational complexity (less storage memory).

Frosst and Hinton presented in their work [4] “Distilling a Neural Network into a soft decision tree” a great approach to explaining DNNs using decision trees. However, their work differs from Aytekin’s paper as they combined the advantages of both DNN and decision trees.

The Crucial Part of the Idea

Building the spanning tree by computing the new weights: the suggested algorithm takes the signals that come to the network and searches for the signals where the ReLUs are activated and where they are not activated. Eventually, the algorithm (transformation) replaces/puts a vector of ones (or the slopes values) and zeros.

The algorithm runs over all the layers. For each layer, it sees what the inputs from the previous layer are and calculates the dependency for each input. Actually, in each layer, a new **effective filter** is selected so it will be applied to the network input (based on the previous decision). By doing so, a fully connected DNN can be represented as a single decision tree where the effective matrix, found by the transformations, acts as categorization rules.

You can also implement it for a convolutional layer. The main difference is that many decisions are made on partial input regions rather than the entire input to the layer.

About **dimensionality and computational complexity**: The number of categories in the obtained decision tree appears to be huge. In a fully balanced tree, we need 2 to the power of the depth of the tree (intractable). However, we also need to remember the violating and redundant rules that provide lossless pruning.

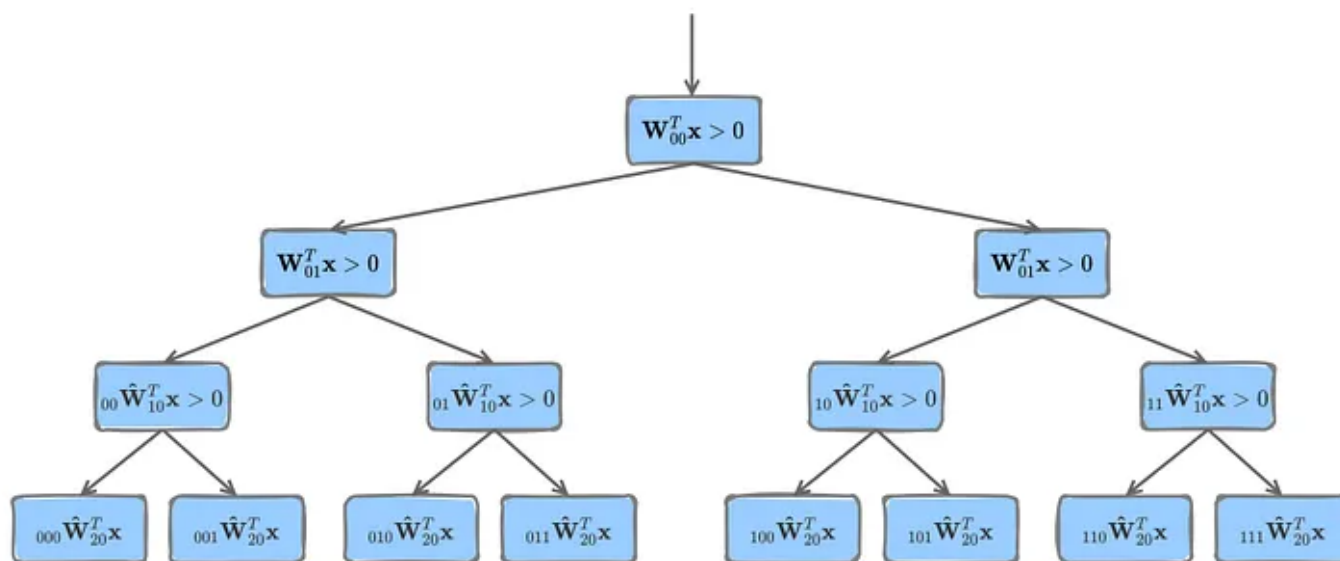


Image by author

Summary and Things to Keep in Mind

- This idea holds for DNN with piece-wise linear activation functions
- The basis of this idea that neural networks are decision trees is not new
- Personally, I found the explanation and mathematical description very straightforward [1], motivated to use it and boost the Explainable AI domain
- Someone needs to test this idea on ResNet 😊

References

The original paper can be found at: <https://arxiv.org/pdf/2210.05189.pdf>

[1] Aytekin, Caglar. "Neural Networks are Decision Trees." *arXiv preprint arXiv:2210.05189* (2022).

If you want to watch a 30 min. interview about the paper look here:

[2] The great Yannic Kilcher interviews Alexander Mattick about this paper, on YouTube: https://www.youtube.com/watch?v=_okxGdHM5b8&ab_channel=YannicKilcher

A great paper on applying approximation theory to deep learning to study how the DNN model organizes the signals in a hierarchical fashion:

[3] Balestrieri, Randall. "A spline theory of deep learning." International Conference on Machine Learning. PMLR, 2018.

A great work that combines the power of both decision trees and DNNs:

[4] Frosst, Nicholas, and Geoffrey Hinton. "Distilling a neural network into a soft decision tree." arXiv preprint arXiv:1711.09784 (2017).

You can read a post on Medium summarizing this work [4]:

[5] Distilling a Neural Network into a soft decision tree by Razorthink Inc, Medium, 2019.

About the Author

Barak Or is an Entrepreneur and AI & navigation expert; Ex-Qualcomm. Barak holds M.Sc. and B.Sc. in Engineering and B.A. in Economics from the Technion. Winner of Gemunder prize. Barak finished his Ph.D. in the fields of AI and Sensor Fusion. Author of several papers and patents. He is the founder and CEO of ALMA Tech. LTD, an AI & advanced navigation company.

Deep Learning

Artificial Intelligence

Machine Learning

Neural Networks

Data Science

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

By signing up, you will create a Medium account if you don't already have one. Review our [Privacy Policy](#) for more information about our privacy practices.



Get this newsletter

[About](#) [Help](#) [Terms](#) [Privacy](#)

Get the Medium app

