Progress in a New Visualization Strategy for ML Models

Alex Wissing
University of Nebraska at Omaha

Brian Ricks
University of Nebraska at Omaha, bricks@unomaha.edu

Robin Gandhi
University of Nebraska at Omaha, rgandhi@unomaha.edu

Yonas Kassa
University of Nebraska at Omaha, ykassa@unomaha.edu

Akshay Kale
University of Nebraska at Omaha, akale@unomaha.edu

Follow this and additional works at: https://digitalcommons.unomaha.edu/isqafacproc

Part of the Databases and Information Systems Commons

Please take our feedback survey at: https://unomaha.az1.qualtrics.com/jfe/form/SV_8cchtFmpDyGfBLE

Recommended Citation
Wissing, Alex; Ricks, Brian; Gandhi, Robin; Kassa, Yonas; and Kale, Akshay, "Progress in a New Visualization Strategy for ML Models" (2023). Information Systems and Quantitative Analysis Faculty Proceedings & Presentations. 63.
https://digitalcommons.unomaha.edu/isqafacproc/63
# Progress in a New Visualization Strategy for ML Models

## A Mid-Project Summary of Design and Ongoing Problems

**University of Nebraska at Omaha, College of IS&T**

---

### Background

#### Overview of ML Models

Machine Learning Models are trained on datasets which are encoded into numbers. The models are then evaluated for effectiveness, eventually yielding a model which can somewhat accurately make predictions based on correlations from the input.

Machine Learning Models take a set of features, or categories of data, and encode them into numbers; a model then is able to be sent information to make predictions. After a prediction is made, it must be interpreted via decoding. This then leads to a result.

#### Visualization

**Main Idea**

This visualization strategy takes two selected features and draws a graph with the model’s output embedded in a color gradient.

At a coordinate (feature 1, feature 2), the model can be ran by generating random, acceptable values for additional features. The model is polled a number of times, each time with the selected features remaining the same, but other features randomly generated. Model output is aggregated in an arithmetic mean value.

This mean value is used as the third dimension in the graph, the color gradient.

### Design

**Main Idea**

This visualization strategy takes two selected features and draws a graph with the model’s output embedded in a color gradient.

At a coordinate (feature 1, feature 2), the model can be ran by generating random, acceptable values for additional features. The model is polled a number of times, each time with the selected features remaining the same, but other features randomly generated. Model output is aggregated in an arithmetic mean value.

This mean value is used as the third dimension in the graph, the color gradient.

---

### Challenges

- Achieving linguistic performance and memory control when models originate from Python-based Tensorflow programs
  - Resolved with the frugally-deep github repository and using C++ to generate visualization data
- Learning about ML and how it’s organized just before and just after a prediction is made
  - Talking with my team members and experimenting with the model in Python and C++ has given me a better understanding of machine learning
- Managing different types of input (Discrete vs Continuous)
  - An incremental approach was used to explore the continuous space, represented using floating-point values. The discrete values are only calculated once if not matched with a continuous, saving runtime

### Work In-progress

- Incremental Variance Algorithm alongside arithmetic mean
- Expanding visualization technique beyond binary models
- Managing inter-feature constraints during random generation and feature selection
- Usage of this visualization technique to explore a model trained on the Iris dataset

## Acknowledgements

This investigation was financially supported by the Army Corps of Engineers.