SYNTHETIC DATA

Synthetic data is information that's artificially manufactured rather than generated by real-world events. Since Deep Neural Networks (DNNs) need large amounts of data for training, synthetic data is particularly useful to train models from a small set of data.

NeuTigers's synthetic data generation toolkit relies on the TUTOR* Deep Neural Networks (DNNs) synthesis framework developed at Princeton University with NeuTigers co-founder Professor Niraj K. Jha's lab.

The data generated by the synthetic data library has the same probability distribution as the real data. The framework validates the integrity of the generated data in real-time.

The framework also benefits from other developments from Princeton University. Mimicking how the human brain works, we optimize the Neural Networks in weights and architecture to reduce its model size while maintaining accuracy by combining the TUTOR DNN framework with the grow-and-prune paradigm.

It is similar to how a toddler interprets the world, humans create energy-efficient "predictive models" that are very accurate, despite using very small data during the learning phase.

The human brain can carry out new tasks with limited experience. It utilizes prior learning experiences to adapt the solution strategy to new domains.

Unlike the human brain, Deep Neural Networks (DNNs), on the other hand, generally need large amounts of data and computational resources for training.

This is where NeuTigers' TUTOR DNN synthesis framework steps in. It synthesizes accurate DNN models with limited available data, reduced memory and computational requirements.

TUTOR produces synthetic data in three steps:

1. Drawing synthetic data from the same probability distribution as the training data and labeling the synthetic data based on a set of rules extracted from the real dataset.
2. Using two training schemes that combine synthetic data and training data to learn DNN weights.
3. Employing a grow-and-prune synthesis paradigm to learn both the weights and the architecture of the DNN to reduce model size while ensuring its accuracy.

NeuTigers synthetic data library reduces the need for labeled data by 5.9% and improves output accuracy by 3.4% despite using a smaller sample size during the learning phase.

Our technology uses fewer samples than Generative Adversarial Networks (GANs), a less effective tool for generating new synthetic data.

*Refer to the following article for details and performance analysis: TUTOR