Contribution ID: **31**   Type: **not specified**

A Novel Approach to Training Foundation Models for Jet-Related Tasks Without Vector Quantization

Thursday 7 November 2024 17:00 (20 minutes)

This study proposes a new method for training foundation models designed explicitly for jet-related tasks. Like those seen in large language models, a foundation model is a pre-trained model that can be fine-tuned for various applications and is not limited to a specific task. Previous approaches often involve randomly masking inputs, such as tracks within a jet, and then predicting the masked parts. However, unlike methods in other fields like image recognition and point clouds, these proposed techniques show less improvement in accuracy for downstream tasks as the amount of training data increases when compared to models trained from scratch.

Most existing methods heavily rely on vector quantization, which is crucial in determining accuracy. In High Energy Physics (HEP), input variables often have highly skewed distributions, making them poorly suited for vector quantization. Additionally, vector quantization using neural networks is known to be very unstable during training.

In response to these challenges, we propose a method that reconstructs masked inputs without using vector quantization. To reduce biases introduced by the model architecture, we use a LLaMA-type Transformer. This approach aims to evaluate the effectiveness of pre-training methods that do not rely on HEP-specific knowledge. We also discuss the results of pre-training and fine-tuning using the JetClass dataset.

Track

Author: MORINAGA, Masahiro (University of Tokyo (JP))

Co-authors: TANAKA, Junichi (University of Tokyo (JP)); SAITO, Masahiko (University of Tokyo (JP))

Presenter: MORINAGA, Masahiro (University of Tokyo (JP))

Session Classification: Foundation models