

A Tree-to-Token Paradigm for Jet Physics: Bridging High-Energy Phenomenology and Language Modeling

Wednesday, 26 February 2025 13:30 (1 hour)

We propose a novel approach for more advanced analysis of jets, which serve as crucial objects of study in high-energy particle physics experiments. While conventional methods often treat jets as point clouds, our work focuses on the binary tree structure obtained during clustering. It explores ways to handle its constituents (such as tracks) using natural language processing language models. Specifically, we convert the binary tree derived from clustering into a bracketed representation, serialize it into a one-dimensional sequence, and then apply tokenization (quantization) to produce a data format suitable for training Transformer models. In this presentation, we will discuss the generation of the tree structure, the tokenization process, and, if time permits, the results of Transformer-based training, thereby demonstrating the potential of this novel perspective for jet analysis.

Presenter: MORINAGA, Masahiro (ICEPP, The University of Tokyo)

Session Classification: Afternoon Talks