Problems Before the Meeting

Ting-Kai Hsu

NOv 20

Contents

1 Where Are the Functions Used

In the block [3] and [4], we can see there are definitions for 2 functions. One is **stack_arrays** and the other is **pad_array**. However, I never see the first one (**stack_arrays**) been used in the remaining part of the code. The latter one has been used at block [5] in the definition of **load()** function. In **load()** function, we append the padded version of array into the new one.

2 An Useful Tool To Understand Awkward Array

Here is the link: click.

3 What is Mask

In the definition of __**init**__ function in **Dataset**, we can see one of the key of feature dictionary would be 'mask'. It confuses me because I don't know the meaning of mask and its function.

Moreover, I don't know why the code doesn't contain any information about the charge carried by the particles.

3.1 Answer to Mask

Mask is a feature of awkward array. It provides the selection to the original array by assigning boolean values to associated array, and they would decide whether the new copy would take that element or not. For more information, please see Awkward Array Release.

4 More on the Needed Information

As we can see in convert_dataset.ipynb, there are many properties for a particle that should be labeled out.

- px, py, pz, e: These would build up the 4 momentum of the particle. **Note that we not only store for jet but also each particle.**
- pt: This would be the transverse momentum, that is, the momentum along the scatter line. **Note that we not only store for jet but also each particle.**
- jet_eta, jet_phi: These two would show the spacial distribution of the particle, which the measurement would along the scattering line.
- jet_mass, n_particle: We store the total mass of jet[1], and also store the number of particles.

I find out we also need to store the relative value[2], which is quite confusing.

[1] I don't know why we need this parameter.

[2] Data of particle relative to jet.

1. part_ptrel: Transverse momentum relative to that of jet, which is

   $$part\_ptrel = \frac{pt}{jet\_pt}$$

   , and also its logirithm value.

2. part_erel: Energy of particle relative to that of jet and its logirithm value.

3. spacial parameters for particle: part_raw_etarel, part_etarel, part_phirel, and part_deltaR.

   - part_raw_etarel: We compute this by formula

     $$\eta_{\text{particle}} - \eta_{\text{jet}}$$

   - part_etarel: This would be the last value multiplied with **eta sign** of jet[3].

     $$\eta_{\text{rel}} = \eta_{\textbf{raw}} \times \text{sign of eta of jet}$$

   - part_phirel: We get this by a function that confuses me. It seems that we can get this paramenter easily.

     ```
             p4.delta_phi(jet_p4)
     ```

   - part_deltaR: This is the spacial distance.

5 Note After Discussion

Professor said that the example provided by author would be merely classification, but eventually we would like to use this model to retrieve the information of B meson[4] The modification we would like to make would be using different loss function. Also, we would like to mention that there is already much more efficient machine learning model in HEP analysis called Particle Transformer. My job now would be try to get some data by simulation, and there is a special simulation package in C++ made for B-tagging, which is called EvtGen. First we can use to classification B+ and B- quark using the model provided without changing

[3] I don't know the meaning of this particle.

[4] This kind of particle can decay in a short time.

the loss function [5]. There is an useful website that contains all possible evolution of elementary particle, called particle data group.

[5] That is, we only need to know how to put the data that is different from the jet tagging in the example into the model.