Contribution ID: **59** Type: **分会报告**

CEPC 上基于 DeepSets 模型的喷注标记算法研究

Monday, 10 July 2023 14:15 (15 minutes)

背景介绍 2012 年在大型强子对撞机（LHC）上发现希格斯（Higgs）玻色子是粒子物理学界的大事件，该发现不仅补全了标准模型缺失的最后一角，而且开启了粒子物理探索的新篇章。由于 *WZH* 三种玻色子（尤其是希格斯玻色子）与新物理现象及新物理规律的联系十分密切，因此精确测量他们的性质，是探索新物理现象和新物理规律的关键手段。强子喷注是 *WZH* 三种玻色子最主要的衰变末态，因此喷注的重建（Clustering）和标记（Tagging）算法对实现 *WZH* 三种玻色子的精确测量至关重要。本文主要介绍喷注的标记算法。传统喷注标记算法有两种，第一种是基于选择条件和人工变量，第二种是基于传统机器学习算法，比如决策树（Boost Decision Tree，BDT）和神经网络（MLP）。近年来，随着深度学习的发展及其性能的进一步提升，越来越多的深度学习算法应用于粒子物理实验中，例如依旧基于人工提取变量方法的 DNN 算法、基于图像识别的 CNN 算法和序列处理的 RNN 算法，以及本文在正负电子对撞机（CEPC）上采取的基于 DeepSets 模型的神经网络方法对夸克喷注标记进行分析。数据、模型和性能本研究采用了新提出的、基于 DeepSets 的能量流深度学习算法以及 *Z* 玻色子强衰变末态产生的喷注模拟数据集，尝试对重味夸克喷注的标记进行研究。其中 *Z* 玻色子衰变为 *bb\bar*，*cc\bar*，(*ss\bar*，*uu\bar*，*dd\bar*) 等 5 种不同强子末态。所有数据通过 Whizard 1.9.5 产生，再用 Pythia 进行强子化模拟，最后再采取 CEPC CDR 的基准探测器对蒙特卡洛样本进行探测器模拟和重建。DeepSets 模型有以下三个特点：第一，喷注中末态粒子的顺序交换不影响喷注的性质，因此 DeepSets 模型必须考虑粒子顺序交换不变性这一对称性；第二，在顺序交换不变性的基础上采用多个的过滤器充分提取喷注的关键物理特征；第三，把提取出的特征作为 DNN 的输入进行标记。以上的数据集和模型在采用 GPU 进行充分训练后，与传统的决策树和 XGboost 算法相比，其性能提高约 6%，使平均精度从 80% 提高到 85% 左右。

Summary

讨论和展望基于 DeepSets 模型的初步研究对夸克喷注标记给出了乐观的结果，下一步作者将进行两方面工作：一是对数据信息进行进一步发掘，以及对模型的超参数进行深入优化，期望得到性能的进一步提高；二是尝试对完整事例进行标记，预期能对 *WZH* 三种玻色子性质的精确测量有较大作用，例如使得 *WZH* 三种玻色子的衰变分支比测量有显著的精度提高。

Primary author: 廖, 立波 (wuzhou univercity)

Co-author: Dr LI GANG(EPD.IHEP), Gang (高能所)

Presenter: 廖, 立波 (wuzhou univercity)

Session Classification: 人工智能与应用

Track Classification: 人工智能与应用