

深度学习在 LHAASO 数据分析中的应用

Tuesday, 16 July 2019 16:10 (20 minutes)

摘要: 高海拔宇宙线观测站 (Large High Altitude Air Shower Observatory, LHAASO) 项目是由中科院高能物理研究所承担, 国内外多家单位联合, 在海拔 4410 米的四川省稻城海子山建立的大型科研项目。LHAASO 是一个复合型探测器阵列, 由 KM2A、WCDA 和 WFCTA 这三个子阵列组成, 将会在 10^{11} eV 到 10^{15} eV 对北天区伽马源进行巡天扫描同时也会对 10^{13} eV 到 10^{17} eV 的宇宙线成分能谱进行测量。一平方公里探测阵列 (KM2A) 是 LHAASO 的主阵列, 主要测量 10^{13} eV 到 10^{15} eV 的高能伽马射线。KM2A 包含电磁粒子探测器阵列 (ED) 和 μ 子探测器阵列 (MD)。其中电磁粒子探测器阵列主要用于记录宇宙线经过广延大气簇射以后产生的次级粒子数密度和到达时间, 用于重建原初宇宙线的方向和能量, μ 子探测器阵列主要测量簇射中的 μ 子信息, 用于成分的鉴别。

LHAASO 实验的核心科学目标是通过伽马射线测量探索高能宇宙线起源并开展相关的高能辐射等研究, 伽马射线和强子宇宙线鉴别是 LHAASO 开展相关科学研究的重要基础和前提。由于探测器得到的数据事例大多数是宇宙线中的带电粒子所贡献的, 所以 KM2A 阵列想要实现探测高能伽马射线源的科学目标, 则需要大量的宇宙线背景中的挑选出伽马射线。所以伽马射线和宇宙线的成分区分能力至关重要, 此前的成分区分方法主要以 μ 子数与电子数的比值作为特征区分和基于多变量分析 (TMVA) 中的机器学习方法, 如随机森林 (Random Forest, RF)、神经网络 (Artificial Neural Network, ANN), 此方法的优点是训练速度快和计算机硬件条件不需要太高, 其缺点是需要人工选取特征和不能提取更深层维度的特征。近年来深度学习飞速发展, 各种新的理论技术不断涌现并逐渐趋于完善, 并且各种新的神经网络结构被提出, 如卷积神经网络 (Convolutional Neural Networks, CNN)、循环神经网络 (Recurrent Neural Network, RNN) 和长短期记忆网络 (Long Short-Term Memory, LSTM) 等, 正在逐渐应用于各个领域。深度学习优势是可以直接基于大量原始信息进行分析和提取更深层维度的特征进而达到精确区分的目的。

本课题将基于谷歌公司开发的开源人工智能框架 TensorFlow, 以 LHAASO 实验装置的 KM2A 的所有探测器记录信息作为输入数据, 对数据进行重建进而实现噪声过滤, 利用光电子数并将 ED 和 MD 阵列图像化, 每个事例的训练数据是两张图片, 分别为 ED 阵列和 MD 阵列的探测器被触发得到光电子值, 再将阵列位置信息添加到模型训练中, 以能量进行划分区间。改进 LeNet 卷积神经网络结构, 使用 ReLu 激活函数, 并使用随机失活 (Dropout) 方法防止模型过拟合等, 最终首次基于深度学习得到了 KM2A 在不同能量处的伽马质子区分能力, 即 Q 因子和 ROC 曲线。另外, 本结果还与传统的方法进行了比较。

关键字: 宇宙线; LHAASO 实验; KM2A 阵列; 深度学习; 卷积神经网络

Primary author: Mr 陈, 秀林 (中科院高能物理研究所)

Co-authors: JIN CHAO; WU SHA (高能所); Mr CHEN (陈), songzhan (松战) (中科院高能物理研究所)

Presenter: Mr 陈, 秀林 (中科院高能物理研究所)

Session Classification: 数据处理与物理软件 I

Track Classification: 数据处理软件与分析方法 II