

Classificação de jatos de partículas em colisões p+p no LHC

Introdução

O LHC, localizado no CERN, é o maior acelerador de partículas do mundo. Ele faz parte de uma série de experimentos que buscam compreender os fundamentos da física de partículas, como a estrutura da matéria e as forças fundamentais da natureza.

Quando prótons são acelerados a altas energias no LHC e colidem entre si, ocorrem interações extremamente energéticas. Nessas colisões, a energia é convertida em uma variedade de partículas subatômicas. Os jatos de partículas são um fenômeno observado nessas colisões de alta energia.

Um jato de partículas é essencialmente um grupo de partículas subatômicas, como prótons, nêutrons, elétrons, mésons e muitos outros, que são produzidos em uma direção semelhante a partir de uma colisão de alta energia. Normalmente são resultantes do processo de fragmentação de quarks (leves ou pesados) e até mesmo glúons, que são espalhados nos instantes iniciais da colisão com elevado momento transversal. Eles carregam informações valiosas sobre as propriedades das partículas que os originaram bem como sobre os processos físicos envolvidos na colisão.

Os jatos de partículas são estudados pelos detectores do LHC, como o ALICE, ATLAS e o CMS. Esses detectores são capazes de identificar e medir as propriedades das partículas que emergem das colisões, permitindo aos cientistas reconstruir os eventos e estudar os fenômenos que ocorrem em escalas subatômicas.

Os estudos com jatos de partículas no LHC têm contribuído significativamente para avanços na compreensão da física de partículas, incluindo a confirmação experimental do bóson de Higgs, a busca por partículas exóticas e a investigação de diversos fenômenos no Modelo Padrão.

O Dataset disponível é resultado de simulações realizadas por Kansal et al (arXiv:2106.11535) em um modelo simplificado de detector. Nesse artigo há detalhes sobre o processo de produção e simulação dos jatos, bem como algumas análises já realizadas envolvendo técnicas de aprendizado de máquina. Há várias versões de dataset mas usaremos a versão 1.0, disponível para download em <https://zenodo.org/records/5502543>. São 3 arquivos .CSV. Um para jatos de glúons, outro para jatos de quarks Top e outro para jatos de quarks leves. Há cerca de 180 mil jatos com $p_T \sim 1$ TeV em cada arquivo e são listadas as 30 partículas mais energéticas em cada jato (somente partículas com $mask = 1$ são reais. Isso é feito para formatar o arquivo em um número fixo de colunas quando temos menos de 30 partículas em um jato). Para cada partícula temos listada informações como a os ângulos relativos em relação à direção do jato e momento relativo ao momento do jato. Mais detalhes podem ser vistos na referência citada acima.

Objetivos

O objetivo principal desse projeto é responder à pergunta: É possível classificar jatos como originados de glúons, quarks leves ou pesados, olhando as distribuições de partículas produzidas nesses processos? Para isso, sugerimos os seguintes objetivos secundários.

1. Utilizar técnicas de pré-processamento de dados para preparar o conjunto de dados de para análise.
2. Análise exploratória de dados: Realizar análises descritivas e visualizações para entender a distribuição das variáveis e identificar possíveis padrões iniciais. Se necessário, utilizar técnicas não supervisionadas para identificar agrupamentos e sub-agrupamentos desses datasets. Se necessário, criar novos features, como tamanho do jato, área efetiva, dispersão das partículas, etc.
3. Aplicar algoritmos de aprendizado supervisionado, como regressão logística, árvores de decisão, SVM (Support Vector Machines) e redes neurais, para tentar classificar os jatos como originados de quarks leves ou pesados ou glúons.
4. Avaliar a performance desses métodos em termos de acurácia e precisão.

Entregáveis

1. Relatório técnico detalhado contendo a descrição dos dados, metodologias utilizadas, resultados das análises, interpretação dos insights obtidos e conclusões.
2. Apresentação oral do projeto destacando os principais resultados, desafios enfrentados e possíveis aplicações práticas dos insights obtidos.

Recursos Necessários

- Artigo contendo a explicação sobre a produção de jatos e a simulação do dataset em <https://arxiv.org/abs/2106.11535>
- Datasets (usar versão 1.0) disponível em <https://zenodo.org/records/5502543>

Equipe

O projeto será realizado por uma equipe de alunos da disciplina de Inteligência Artificial, com orientação e supervisão dos professores responsáveis pela disciplina.

Considerações Finais

Este projeto oferece uma oportunidade valiosa para os alunos aplicarem seus conhecimentos em inteligência artificial em um contexto de física de partículas elementares em situações similares àquelas presentes nos experimentos do LHC.