

Использование методов машинного обучения для разделения нейтральных частиц в калориметре LHCb

Междисциплинарный семинар "Математика, компьютерные
науки и информационные технологии"

Виктория Чекалина(ШАД, Яндекс), Фёдор
Ратников(Lambda, ФКН), Александр Панин(Яндекс)

Школа анализа данных

12 февраля 2017 г.



SCHOOL OF DATA ANALYSIS

Задачи Большого Адронного Коллайдера: уточнение существующей модели физики частиц

Постулаты Стандартной модели

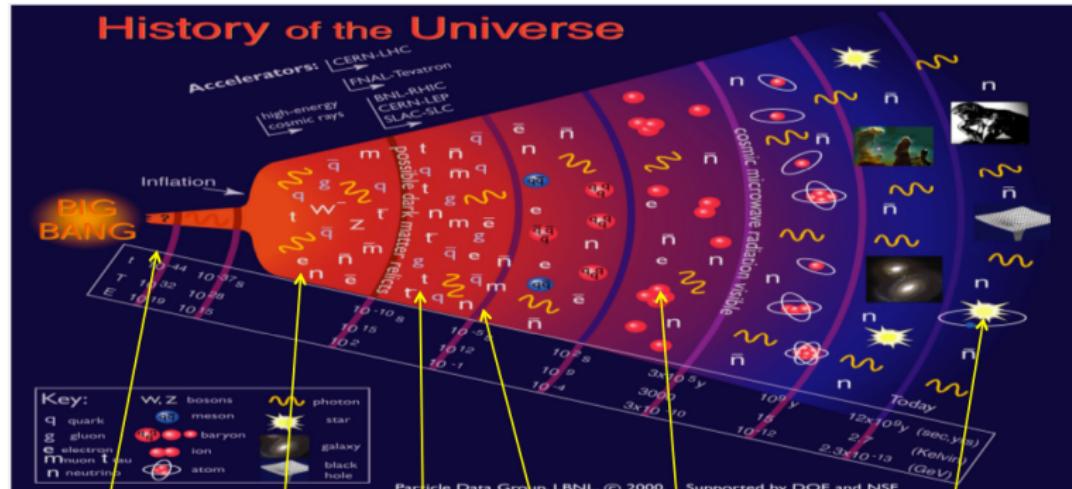
- Вещество состоит из 12-ти элементарных частиц: 6 лептонов и 6 кварков
- Также существуют частицы-переносчики взаимодействия (фотон, глюон, W и Z-бозон)
- Есть 3 типа взаимодействия:
 - Сильное (взаимодействие между нуклонами в ядре)
 - Слабое (β - распад)
 - Электромагнитное (притяжение электронов к ядру)

Вопросы

- А как же Гравитационное взаимодействие?
- При высоких температурах Электромагнитное и Слабое взаимодействие объединяется в Электрослабое, каковы его свойства?



Задачи Большого Адронного Коллайдера: назад во времени



Космология

Физика частиц

Ядерная физика

Космическое излучение

Кварк-глюонная плазма

Астрофизика

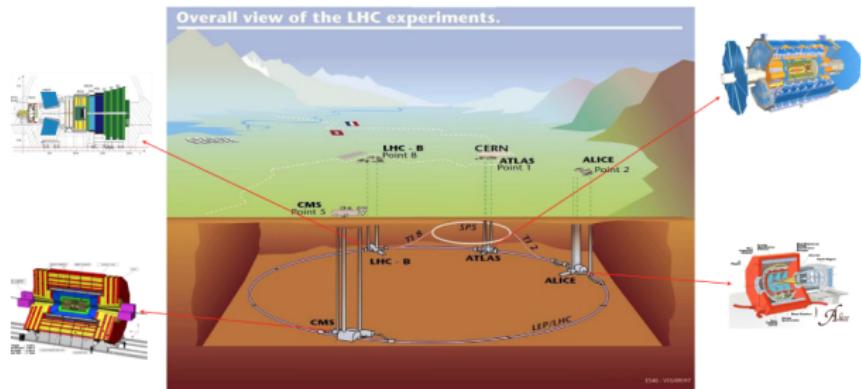
Задачи экспериментов Большого Адронного Коллайдера

- Создать экспериментальную базу для универсальной модели и подтвердить существующую Стандартную модель
- “Заглянуть” во времена рождения Вселенной

Для решения этих задач нужно изучать материю на экстремально малых расстояниях и при экстремально больших температурах. Это можно достичь, сталкивая пучки протонов высоких энергий. При этом действительно интересные вещи наблюдаются очень редко: чтобы поймать их, нужно уметь обрабатывать большое число столкновений в секунду.

Большой Адронный Коллайдер

- Радиус 4 km
- Энергия столкновения 13,6 ТэВ
- Количество столкновений в секунду 1 billion



В большом
адронном коллайдере - 4 больших детекторных комплекса.
Каждый из них заточен для решения определённых задач

Законы сохранения

- С каждым свойством частицы связан закон сохранения или определённая симметрия: с зарядом - обычным Q или барионным B - закон сохранения заряда, с изоспином I - закон сохранения изоспина. В некоторых слабых взаимодействиях законы сохранения нарушаются.

Чётность

- C - четность - неизменность системы относительно зарядового сопряжения
- P - чётность - неизменность системы относительно пространственного отражения
- Комбинированная чётность или CP -инвариантность - неизменность относительно замены частиц на античастицы и одновременного отражения.

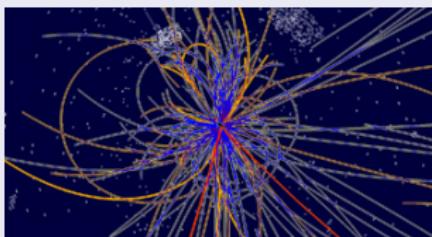
Эксперимент LHCb

- Доказано, что С и Р-чётности могут нарушаться.
- СР-инвариантность также может быть нарушена.
- Изучение свойств СР-нарушения является одной из задач LHCb
- Нарушение СР-чётности означает, что физические законы могут по-разному действовать на частицы и античастицы. Это поможет объяснить, почему после Большого Взрыва материя преобладает над antimатерией.

- Нарушение СР-инвариантности детектируется в процессах распада В-мезонов
- В-мезоны живут 10^{-12} сек. Мы можем детектировать только долгоживущие продукты их распада.

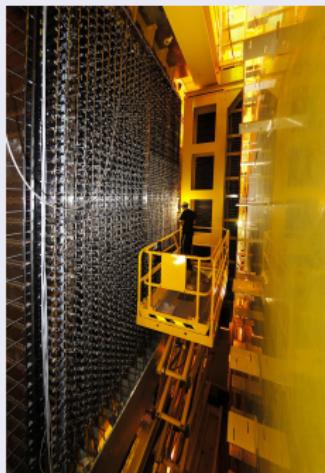
Детектирование заряженных частиц

Заряженные частицы детектируются трековыми детекторами. Трековый детектор, например, может быть пластиной полупроводника, на поверхности которой при пролёте заряженной частицы индуцируется заряд. Несколько таких следов собираются в траекторию частицы, по которой можно определить её тип. Перед трековым детектором обычно располагается магнит: траектории частиц различного заряда и массы по-разному изменяются под действием магнитного поля.



Детектирование незаряженных частиц

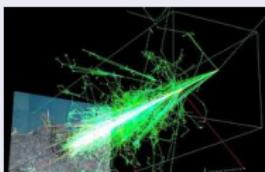
Электро-магнитный калориметр



Незаряженные частицы детектируются в калориметрах. Калориметр - чередующиеся пластины вещества. Пролетая сквозь них, частица распадается на ливень вторичных частиц, которые поглощаются сцинциллятором с последующим испусканием им световых фотонов.

Калориметр показывает, что в данной области пространства выделилась определённая энергия, по которой и классифицируют частицу.

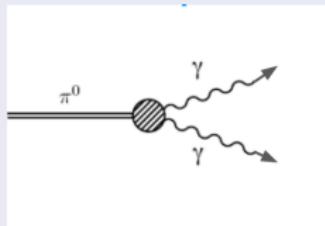
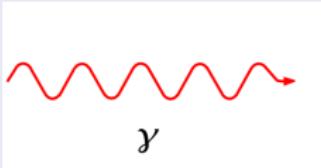
Развитие ливня



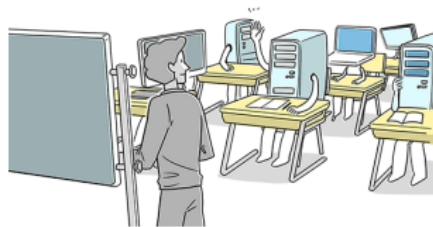
Разделение частиц одинаковой энергии

Что делать, если две близко летящие частицы попадают в одну область и имеют энергию, как одиночная?

Нейтральных пион (π^0) - один из продуктов распада В-мезона - распадается на два фотона. Если распад произошёл вблизи плоскости калориметра, трудно определить, что же всё-таки прилетело.



Машинное обучение



Обучение с учителем

- Имеется множество объектов (у нас - отклик калориметра на частицу) и множество возможных ответов (у нас: фотон или двойной пион). Существует некоторая зависимость между ответами и объектами, но она неизвестна.
- На основании известных пар “объект-ответ” требуется построить алгоритм, способный для любого объекта выдать достаточно точный ответ.
- Для оценки алгоритма используют метрику качества. В нашей задаче используется метрика ROC-AUC - площадь под ROC-кривой.

Машинное обучение

Каждый объект выборки нужно описать, используя ограниченное количество признаков. Нужно выбрать такой набор, по которому объекты наиболее разделимы в данной задаче.

Пример

Задача: отличить яблоко от морковки.

- Хорошие признаки:

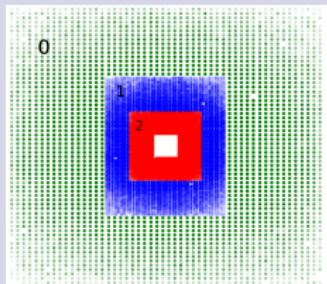
объект\признак	цвет	форма	есть ботва?
яблоко	красный	сферическая	нет
морковь	оранжевый	продолговатая	есть

- Плохие признаки:

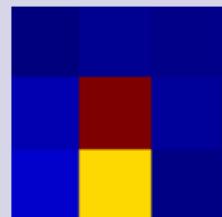
объект\признак	вес	продаётся в магазине?	плод ли это?
яблоко	150 г.	да	да
морковь	150 г.	да	да

Возможное решение задачи

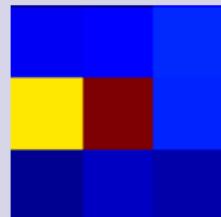
Калориметр



Отклик для
фотона



Отклик для
 π_0



N - размер квадрата, x_c , y_c - центр квадрата

$$S_{xx} = \frac{\sum_{i=1}^N e_i (x_i - x_c)^2}{\sum_{i=1}^N e_i} \quad S_{yy} = \frac{\sum_{i=1}^N e_i (y_i - y_c)^2}{\sum_{i=1}^N e_i}$$

$$S_{xy} = S_{yx} = \frac{\sum_{i=1}^N e_i (x_i - x_c)(y_i - y_c)}{\sum_{i=1}^N e_i}$$

Возможное решение задачи

Признаки - мера асимметрии отклика

- $r2 = \langle r2 \rangle = S_{xx} + S_{yy}$

- $r2r4 = 1 - \frac{\langle r2 \rangle}{\langle r4 \rangle}$

- $k = \sqrt{\left(1 - 4 \frac{S_{xx}S_{yy} - S_{xy}^2}{(S_{xx} + S_{yy})^2}\right)}$

- $asym = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$

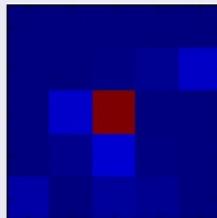
- $\frac{E_{seed}}{E_{cl}}$

- $\frac{E_{seed} + E_{2nd}}{E_{cl}}$

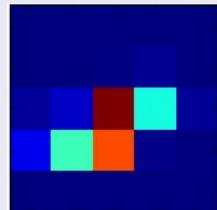
- Классификатор - многослойный перцептрон
- Качество 0.75

Другое возможное решение задачи

Отклик для фотона



Отклик для π_0



Признаки - интенсивность в каждой точке отклика с учётом положения

Классификатор: Neural Network (качество 0.92)

Классификатор: XGboost (деревья) (качество 0.95)

Спасибо!