

ОПРЕДЕЛЕНИЕ БАЗОНА ХИГГСА С ПОМОЩЬЮ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Шидловский О.В.

Смирнова Г.Ф. - к. ф.-м.н. наук, доцент

В Большом адронном коллайдере, протонные пучки ускоряются по круговой траектории в обоих направлениях. Когда эти сгустки пересекаются в детекторе ATLAS, некоторые из протонов сталкиваются, производя сотни миллионы протон-протонных столкновений в секунду. До сотен частиц, вытекающих из каждой сгустки пересечения (так называемый события) обнаруживаются с помощью датчиков, создавая векторы - которых около ста тысяч.

В этих исходных данных рассматриваются: тип, энергия, и 3D-направлении каждой частицы, а также большое количество вещественных параметров задающих частицы.

Все события делятся на интересные и не интересные для нас. Интересные – это те частицы, которые выделяются в процессе столкновения. Не интересные, те которые появляются в виде шума (искажения реальных данных)

Выбранные события записываются на диск, для этого требуется мощные ресурсы и CPU. За год накапливается примерно петабайт данных. Цель работы отличить событие, которое является шумом от события которое реально произошло.

Два протона сталкиваются примерно раз в 50 наносекунд и производят небольшой фейерверк, в котором часть кинетической энергии протонов преобразуется в новую частицу. Большинство полученных частиц очень нестабильны и быстро распадаются в каскад – более легкие частицы. Детектор ATLAS измеряет три свойства этих выживших частиц (так называемый конечное состояние): тип частицы (электрона, фотона, мюонов и т.д.), его энергии, и 3D направление частицы.

Все данные делятся на два типа. Первые тип – это те данные для которых мы уже знаем ответ (было ли событие или это был шум), в дальнейшем будем называть этот тип данных “Тренировочной выборкой”. Второй тип данных – это данные для которых не известно к какому классу принадлежат события. Для каждого события столкновения мы имеем большое количество параметров (одни из них):

- Расчетная масса m_H , полученные через вероятностные интеграции фазового пространство (может быть не определено, если топология случая слишком далеко от ожидаемого топологии)
- Поперечная масса между потерянной поперечной энергии и лептоном.
- Инвариантная масса адронной τ и лептоном.
- Модуль вектора суммы поперечного импульса адронной τ и лептоном
- Абсолютное значение разделения псевдобыстро между двумя струями

Трудности изучения протон-протонных столкновений.

В изучении протон-протонных столкновений есть две главные трудности: одна экспериментальная и одна теоретическая.

- В каждом столкновении рождается слишком много частиц. Некоторые из них при этом вообще не попадают в детектор, а «улетают в трубу», так что разобраться в этой мешанине очень трудно.
- Теоретики умеют хорошо рассчитывать процессы с отдельными кварками или глюонами, но описать адронизацию из первых принципов пока не удается. Адронизацию приходится учитывать с помощью численного моделирования, и поэтому связь между теорией и экспериментом не столь непосредственна, как, например, в электрон-позитронных столкновениях.

Из за трудностей распознавания столкновений используя машинные методы анализа (алгоритмы машинного обучения) которые позволяют создать модель обученную на большом наборе данных.

Так как мы имеем около 350000 событий для которых уже известен ответ, построим модели машинного обучения для их классификации.

Рассмотрим 3 основных приема:

- Классификация с помощью “Дерева решений”
- Классификация с помощью “Нейронных сетей”
- Классификация с помощью “Метода опорных векторов”

Метод “Дерева решений”: для решения поставленной задачи с помощью этого метода определим функцию, которая показывает насколько правильно был получен результат.

Для построение модели позволяющей точно определять класс события построим бинарное дерево. Которое в каждом узле будет содержать условие перехода в левую или правую часть. Для выбора такого условия и коэффициентов, будем перебирать значение функции погрешности до тех пор пока она не станет минимальной. После чего строим под деревья таким же образом пока не придем к окончательному ответу (погрешность = 0). В итоге мы получим бинарное дерево, которая может классифицировать новое событие - пройдя от корня дерева до листа выполняя условия в узлах.

Недостаток: медленная работа алгоритма.
Плюсы: Хорошая точность.

Метод “Нейронных сетей”: для решения поставленной задачи с помощью этого метода построим математическую модель, которая будет состоять из нескольких так называемых слоев. Каждый слой соединен с другим дугами и представляет собой граф состоящий из нескольких соединенных вместе двудольных графов. На первый слой на каждый узел подается информация (в нашем случае свойства каждого события) затем переходя от одного слоя к другому мы получаем в последнем слое коэффициенты задающие правильный ответ. Если ответ не верные мы проходим в обратном направлении и с помощью функции “обратного распределения” меняем веса ребер. После выполнения большого числа операций мы получаем “на тренированную” модель в которую можно подавать новые данные на первый слой и получать ответ.

Недостаток: Плохая точность при плохой выборки данных
Плюсы: Быстрый и точный алгоритм (в случае хорошей выборки тренировочных данных)

Метод “Опорных векторов”: основная идея метода — перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей наши классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора.

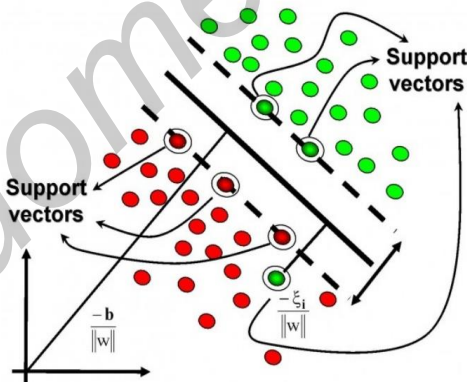


Рис. 1. Иллюстрация принципа работы метода опорных векторов

Список использованных источников:

1. Сегаран. Т. Програмируем коллективный разум. – Пер. с англ. – СПб: Символ-Плюс, 2008.
2. Интернет ресурс - <http://www.kaggle.com/c/higgs-boson>