Использование методов машинного обучения для поиска оптимальных конфигураций детектирующих систем

Семинар "Использование новых методов обработки данных физического эксперимента. Применение методов машинного обучения на комплексе NICA"

Курбатов Евгений[1,2]
Ратников Федор[1,2]

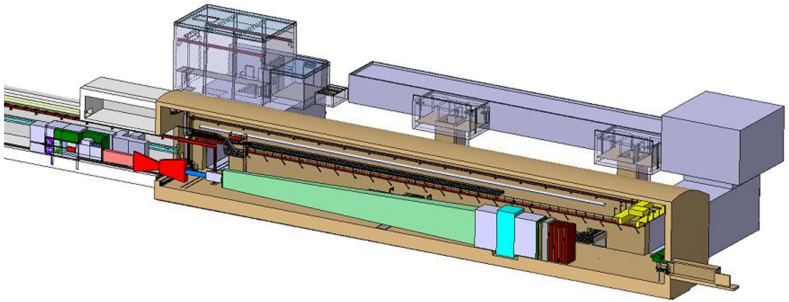
[1]НИУ ВШЭ
[2]Yandex School of Data Analysis

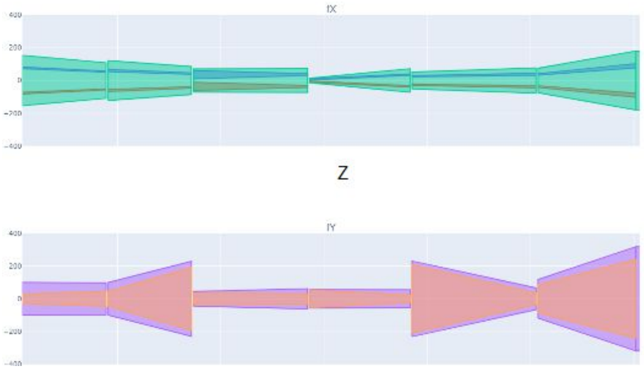
Санкт-Петербург, 2023

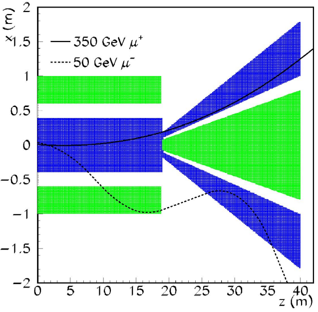
Обзор проблематики

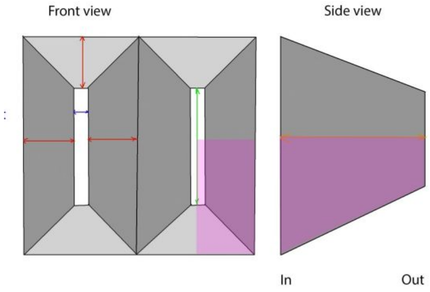
- В настоящее время одним из актуальных направлений применения машинного обучения в физике высоких энергий являются задачи поиска оптимальных конфигураций детектирующих систем. Целью подобной оптимизации является нахождение баланса между способностью всех детекторов бесконфликтно выполнять поставленные задачи и стоимостью постройки, или модернизации установки.
- В данной работе рассказывается о подходах к комплексной оптимизации с применение методов машинного обучения детекторных систем в сложных экспериментах на примере оптимизации мюонной защиты эксперимента SHiP. Основными факторами успеха оптимизации являются корректный выбор целевой функции, метода оптимизации и способа быстрой оценки конфигураций.
- В докладе будут обсуждены проблемы выбора целевой функции, учет ее ограничений с точки зрения эксперимента. Представлены различные подходы к глобальной оптимизации, приемы ускорения расчетной компоненты задачи.
- Приведены результаты применения описанных подходов в эксперименте SHiP.

Обзор эксперимента SHiP









arXiv:1703.03612

- SHiP - новый эксперимент, планируемы на ускорителе CERN SPS
- Один из ключевых элементов - мюонная защита, магнитная система длинной ~30м
- Конфигурация защиты описывается через 42 параметра
- Качество защиты - предмет дискуссии нескольких научных подгрупп, отвечающих за ряд систем детектора
- Подавление мюонов ~6 порядков из 10^{11} на спилл
- Прохождение мюонов через защиту - процесс стохастический

Типичный симуляционный сетап эксперимента

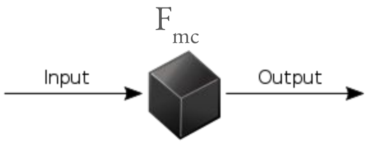
Неотъемлемой частью любого крупного эксперимента в сфере физики высоких энергий на текущий день является наличие модели компьютерной симуляции основных его частей. Как правило, в подобную симуляцию включены:

- Сэмпл/генератор входящих частиц (PYTHIA, G4 Particle Gun, G4 GPS, e.t.c.)
- Модуль симуляции прохождения частиц через сконфигурированный объем детектора (GEANT 3/4, FLUKA, e.t.c.)
- Преобразование энерговыделения частиц в “хиты” внутри чувствительных объемов детекторов
- Преобразование “хитов” в отклик детектора
- Преобразование “хитов” в треки частиц и дальнейшая обработка

В дальнейшем предполагается, что подобная модель существует, работает и описывает установку с достаточным уровнем точности и детализации. Также предполагается, что существует некоторый набор параметров, который отвечает за форму/размер/положение/число элементов детектора, которые можно варьировать в некотором диапазоне.

Постановка задачи

Предположим, что у нас есть возможность проводить симуляцию работы детектора методами, описанными на прошлом слайде. Будем рассматривать такую симуляцию как black box функцию F_{mc}.



Результатом работы симуляции будет некоторый результат F_{mc}(input_sample, p_i), где p_i - это варьируемые параметры детектора, а input_sample - набор частиц на входе в детектор.

Считаем, что у научной группы есть некоторые представления о том, как F_{mc}(input_sample, p_i) характеризует качество использованной для симуляции конфигурации детектора, которые выражаются в наборе численных (зачастую противоречащих друг другу и имеющих различную значимость) метрик M_i.

Задачей оптимизации будет нахождение такого набора p_i, который даст наилучший результат F_{mc}(input_sample, p_i) с точки зрения полученных M_i, с учетом их важности для эксперимента.

Оптимизация = минимизация

Глобальная оптимизация - это область прикладной математики и численного анализа, в которой предпринимаются попытки найти глобальные минимумы (максимумы) функции или набора функций на заданном множестве. Нахождение глобального минимума функции гораздо сложнее нахождения локального: аналитические методы часто оказываются неприменимыми, а использование численных стратегий решения часто приводит к очень сложным задачам.

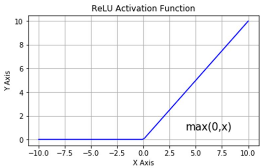
С практической точки зрения наиболее разумно рассматривать оптимизацию с одномерным критерием оптимальности.

Таким образом, одним из наиболее важных шагов для оптимизации сложного детектора является построение целевой функции $F \in \mathbb{R}$ на основе описанных ранее критериев работы детектора M_i.

Принципы построения целевой функции

Хотя построение конкретной функции F является творческим процессом, завязанным на индивидуальные детали каждого отдельного случая, можно выделить ряд свойств и подходов, на которые желательно обратить внимание вне зависимости от реализации алгоритма оптимизации.

- Избегайте исчезающих градиентов
- Приведите вклады M_i в целевую функцию к масштабу, соответствующем их важности
- Если у какой-то из метрик есть которой нет смысла улучшаться - ис или похожую функцию

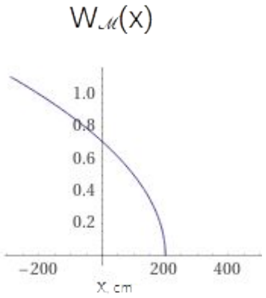


- Старайтесь использовать информацию, которая может и не улучшать M_i напрямую, но потенциально свидетельствовать о движении в нужном направлении. Особенно актуально для дискретных M_i.
- Если имеющейся информации из симуляции недостаточно, есть смысл добавить новые чувствительные объемы

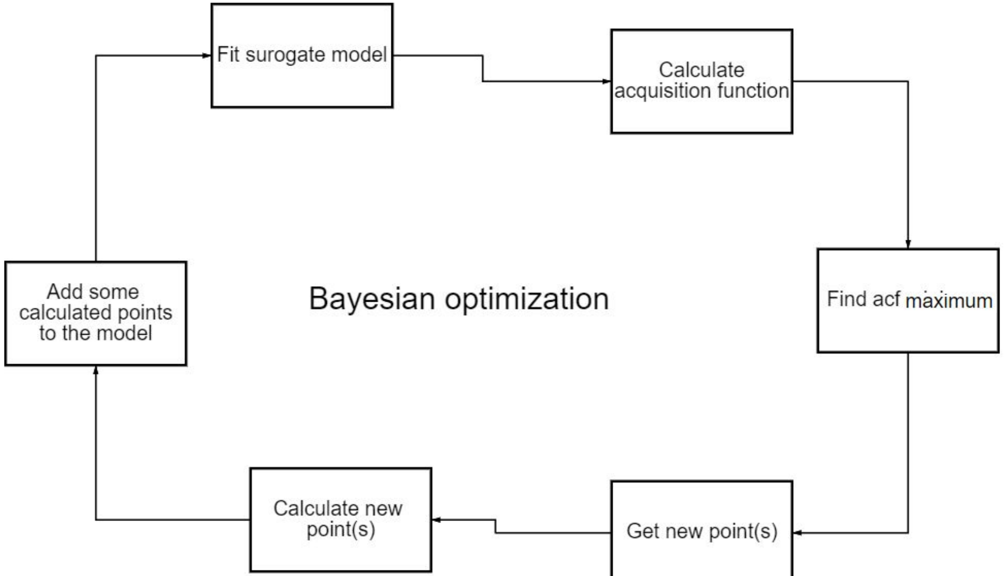
Пример целевой функции для оптимизации мюонной защиты в эксперименте SHiP

$$(1 + e^{\frac{10 \cdot (M - M^*)}{M^*}}) * (1 + \Sigma W_\mu)$$

M -масса защиты, M* - константа, W_μ функция от координаты хита мюона



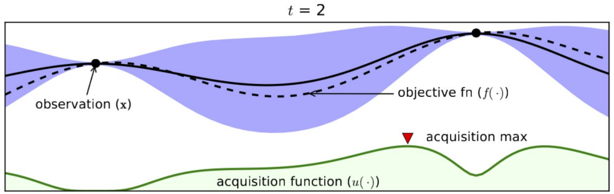
Байесовская оптимизация



Одним из наиболее распространенных методов оптимизации black box функций, чье вычисление требует значительных ресурсов, является Байесовская оптимизация.

Ее использование в базовом варианте оправдано, если:

- Размерность пространства параметров < 10
- Целевая функция может рассматриваться как непрерывная
- Форма пространства параметров не является слишком сложной
- Информация о градиенте целевой функции недоступна



Байесовская оптимизация. Суррогаты и размерность

Базовым и наиболее распространенным суррогатом для Байесовской оптимизации являются Гауссовы Процессы.

Задачи глобальной оптимизации сложных детекторов являются задачами высокой размерности, что влечет за собой большое необходимое число проверенных конфигураций. Из-за этого на передний план выходят 2 главных недостатка подхода:

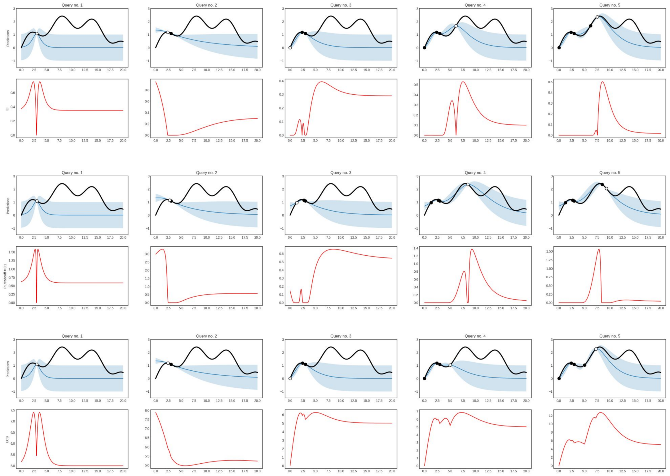
- вычислительная сложность $O(n^3)$
- требования к памяти $O(n^2)$

Это приводит к поиску способов повысить скорость работы, снизить требования к памяти, или уменьшить необходимое число итераций.

Наиболее удачно показала себя комбинация:

- Старт с набора случайных точек
- High-dimensional Bayesian optimization with sparse axis-aligned subspaces (PMLR 161:493-503, 2021)
- Variational Nearest Neighbor Gaussian Process (https://arxiv.org/abs/2202.01694)

Байесовская оптимизация. ACQF.



Expected Improvement:
EI(x) = E(max(f(x) - best_f, 0))

Probability of Improvement:
qPI(X) = P(max Y >= best_f), Y ~ f(X), X = (x_1,…,x_q)

Upper Confidence Bound
UCB(x) = mu(x) + sqrt(beta) * sigma(x)

В программных пакетах, как правило, представлены данные разновидности функций с вариациями на случай наличия шумов при расчете целевой метрики.

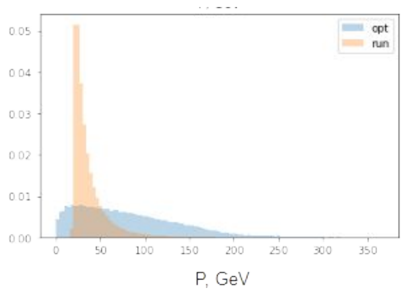
Сэмпл. Форма. Скорость.

Большое число конфигураций, для которых необходимо провести симуляции накладывает ограничения на число событий, которые могут быть обработаны в рамках одной симуляции.

С одной стороны, слишком большое число событий существенно замедлит процесс оптимизации, т.к. больше времени будет уходить на каждую итерацию.

С другой стороны, при недостаточном числе событий, полученные оценки для целевой функции могут иметь слишком большую неопределенность.

Хорошо себя показал метод динамической подстройки: в случае нехватки данных для оценки целевой функции с заданной точностью запускалась повторная итерация симуляции с дальнейшим усреднением результатов, до достижения нужной точности.



Стоит отметить, что выбор состава сэмпла, а также формы импульсных распределений в значительной степень влияет как на результат оптимизации, так и на скорость симуляции.

Сэмпл. Размер. Веса.

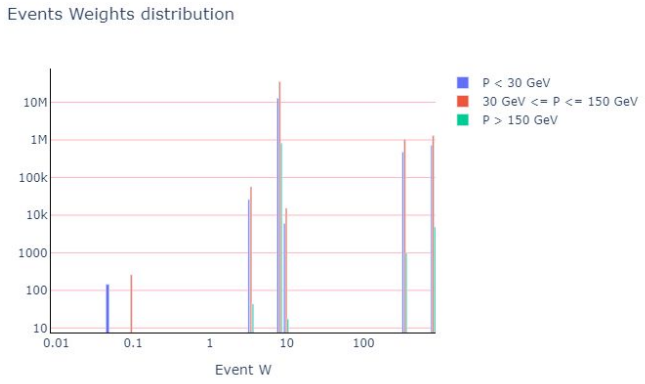
Еще одной проблемой работы с сэмплами с низким числом событий является проблема, возникающая при использовании в расчетах метрик весов этапа генерации.

Если таковые используются, необходимо убедиться, что они не дестабилизируют результат.

Если такая проблема существует, то хорошим способом ее решить будет переход к невзвешенному сэмплу через процедуру “раскрутки” весов:

- из исходного сэмпла выбирается случайным образом выбирается событие (с вероятностью пропорциональной его весу) и добавляется в новый сэмпл
- в исходном сэмпле вес этого события уменьшается на минимальный из имеющихся в сэмпле весов
- процедура повторяется до достижения нужного числа событий в новом сэмпле

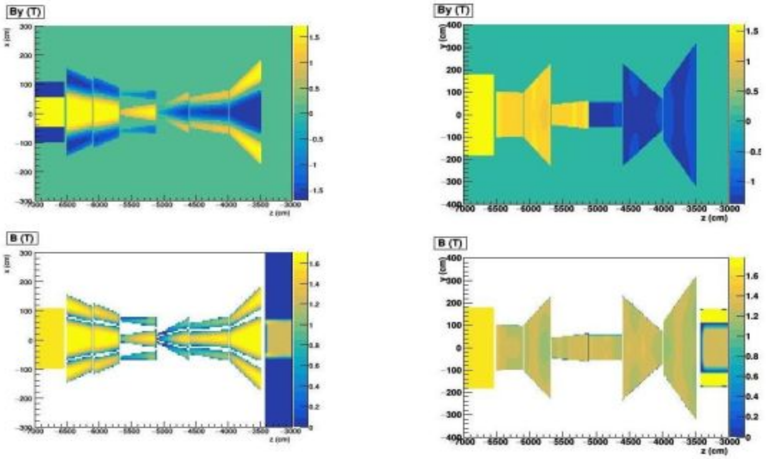
Данный подход позволил значительно снизить уровень неопределенности при работе с сэмплами ограниченных размеров.



Магнитные поля и их карты

- Для увеличения точности симуляции детектора в крупных экспериментах стараются использовать реалистичные карты магнитного поля для больших объемов, где поле потенциально может быть неоднородным.
- Создание такой карты для конкретной геометрии как правило занимает значительное время, что препятствует использованию таких карт в симуляции при оптимизации, если объемы, к которым привязаны карты изменяются.
- Игнорирование разницы наличия и отсутствия карты полей делает результаты оптимизации менее релевантными.
- Наиболее удачно показал подход с созданием синтетической карты поля:
    - В первый запуск симуляции производится сканирование нужной области и построение “идеальной” карты поля
    - “Идеальная” карта модифицируется с целью придать ей более реалистичный вид
    - “Боевой” запуск симуляции производится уже с модифицированной картой поля

Примеры искусственных карт магнитных полей для оптимизации мюонной защиты эксперимента SHiP. Сверху использовалась гауссово размытие + обрезка по форме объема для имитации краевых эффектов. Снизу - внедрение неоднородностей в поле с целью повышения устойчивости полученных конфигураций.



Результаты оптимизации в SHiP

Shield configuration	Tracker rate [Muons/spill]	Shield length [m]
ECN4	45k	31
ECN3 Combi	160k	26
ECN3 Combi Optimized	67k	26
ECN3 SC Optimized	23k	21

Представлены 2 независимых цикла оптимизации:

- Сокращение длины защиты на 5м
- Сокращение длины защиты еще на 5м и замена первых трех магнитов одним сверхпроводящим

ECN4 - изначальная конфигурация защиты
ECN3 Combi - попытка ручной адаптации

Выводы

- Рассмотрена проблема глобальной оптимизации сложных детекторов
- Предложено решение проблемы на основе Байесовской оптимизации
- Выделены основные аспекты построения целевой функции
- Предложены решения, повышающие скорость оптимизации и стабильность найденных конфигураций

Софт

Ax ⟷ BoTorch ⟷ GPyTorch

Верхний уровень. Сохранение/загрузка моделей, подготовка данных, настройка пайплайна

Средний уровень. Все, что касается Байесовской оптимизации, кроме суррогата.

Все что касается суррогатов - тут

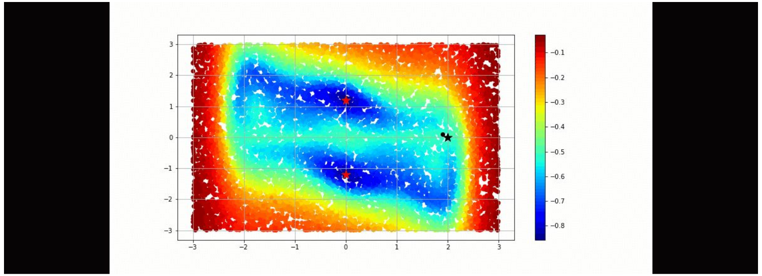
Local Generative Surrogates

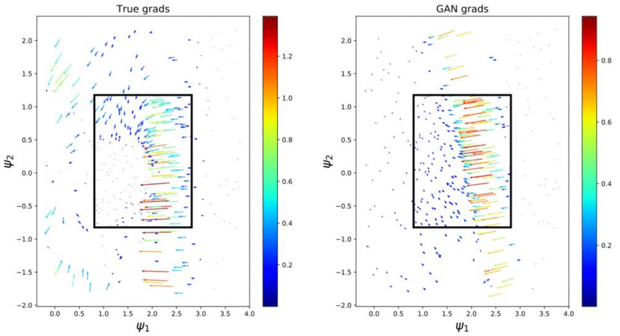
Как можно было увидеть из презентации, байесовская оптимизация не ставит себе задачу попасть точно в минимум, в задачах большой размерности это редко достижимо.

Градиентный спуск и его разновидности - хороший способ для достижения локального минимума.

Но через МС симуляцию нельзя прокинуть градиент.

Если получить достаточно хорошую генеративную модель, предсказывающую результаты симуляции, то через нее можно прокинуть градиент!





https://arxiv.org/abs/2002.04632