

Доменное состязательное обучение для понижения
смещения прогноза при поиске бозона Хиггса в детекторе
ATLAS

Ильяс Фатхуллин

Московский физико-технический институт

26.11.2019

Задача классификации событий в детекторе ATLAS

Выборка: получена из генератора, который моделирует процессы $ttH(H \rightarrow bb)$ и $tt + bb$. Элементы выборки (x_i, y_i) будем называть событиями. Каждое событие характеризуется вектором физических величин x_i , характеризующих теоретическое представление о данном процессе.

Цель: предложить метод классификации событий $G : \mathbb{X} \rightarrow \mathbb{Y}$, которые соответствуют процессам производства $ttH(H \rightarrow bb)$ и $tt + bb$.

Проблема физического эксперимента: распределение \mathcal{D} случайных величин, реализации которых представлены данной выборкой $\{(x_i, y_i)\}_{i=1}^m$, несколько отличается от реальных данных, что приводит к увеличению ошибки классификации \mathcal{L} .

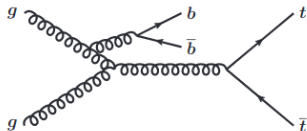
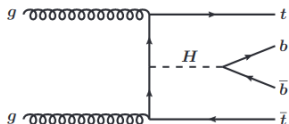
Проблема построения модели: модель, которая использует данные только из одного распределения показывает низкое качество классификации.

Задача классификации событий в детекторе ATLAS

Задача: построить алгоритм бинарной классификации

$G(\cdot) = G_y(G_f(\cdot)) : \mathbb{X} \rightarrow \mathbb{Y}$, $\mathbb{Y} = \{\text{bkg}, \text{sgnl}\}$, который

- понижает смещение прогноза σ_b ,
- показывает адекватное качество (точность, AUCROC) на данных из альтернативного генератора,
- обобщается на реальные данные,
- обобщается на случай данных из нескольких генераторов.



Диаграммы Фейнмана для процессов ttH (сигнал, слева) и $tt + bb$ (фон, справа)

Сигнал отличается от фона тем, что в этом процессе производится бозон Хиггса, с последующим распадом на два нижних кварка. Партонные начальные и конечные состояния у обоих процессов одинаковые, отличие может быть выявлено лишь из кинематики.

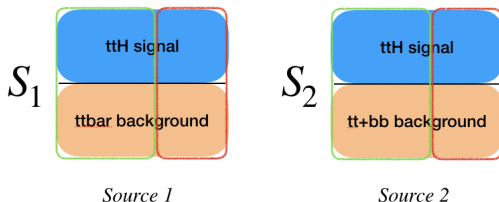
Генераторы данных

Два набора данных симулированных S_1 и S_2 получены из генераторов:

- ttH signal: MadGraph/Herwig6,
- ttbar background: MadGraph/Pythia6,
- tt+bb background: Powheg Pythia8

\mathbb{X} пространство признаков ($x \in \mathbb{X}$, $\dim(x) = 40$), которые характеризуют отклик детектора ATLAS

\mathbb{Y} множество классов ($y = 0$ если процесс фоновый, $y = 1$ если в процессе рождается бозон Хиггса)



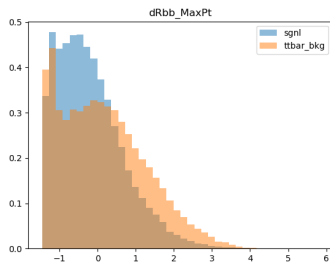
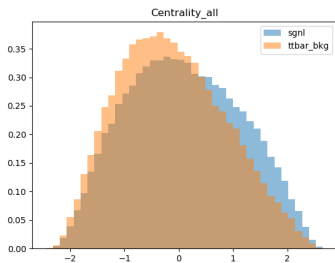
Два набора данных. Отличие в фоновых событиях.

Описание признаков

Num	Name	Description
1	nJets	number of jets
2	MET	missing transverse energy
3	Mjj_MaxPt	mass of jet pair with largest transverse momentum
4	nJets_Pt40	number of jets above 40 GeV p_T
5	nbTag	number of b-tagged jets
6	nHiggsbb30	number of Higgs boson candidates from b-jets, in 30 GeV window
7	nHiggsjj30	number of Higgs boson candidates from all jets, in 30 GeV window
8	pT_jet5	transverse momentum of 5th leading jet
9	pT_lep	transverse momentum of charged lepton
10	HT_jets	sum of transverse momentum of jets
11-14	Hi_all	ith Fox-Wolfram moment, $i=0,1,2,3$
15	H2_jets	2nd Fox-Wolfram moment (jets only)

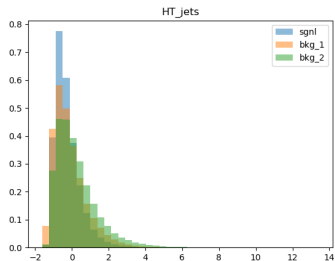
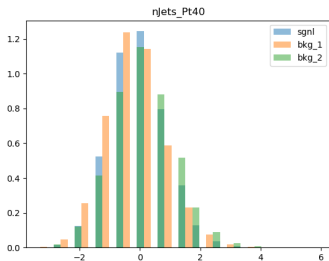
Описание первых 15-ти признаков, которые характеризуют отклик детектора на события

Распределение признаков сигнал/фон1



Распределение признаков Centrality_all (слева) и dRbb_MaxPt (справа).
Распределение сигнальных (синий) и фоновых (красный) событий достаточно близки.)

Распределение некоторых признаков сигнал/фон1/фон2



Распределение признаков для nJets_Pt40 (слева) и dRlb3 (справа).
Различие между распределениями двух фонов того же порядка как и
различие между фоном и сигналом.

sgnl – ttH сигнал, bkg_1 – ttbar фон, bkg_2 – ttbb фон.

Модель классификации

$$y = G_y(G_f(\mathbf{x}, \theta_f, b_f), \theta_y, b_y),$$

$$G_f(\mathbf{x}) = G_f(\mathbf{x}, \theta_f, b_f) = \text{ReLU}(\theta_f^T \mathbf{x} + b_f),$$

$$G_y(G_f(\mathbf{x}), \theta_y, b_y) = \text{softmax}(\theta_y^T G_f(\mathbf{x}) + b_y),$$

$$\text{softmax}(\mathbf{a}) = \left[\frac{\exp(a_i)}{\sum_{i=1}^{|\mathbf{a}|} \exp(a_i)} \right]_{i=1}^{|\mathbf{a}|},$$

где $G_y(\cdot, \theta_y)$, $G_f(\cdot, \theta_f)$ – предсказатель метки и генератор признаков с соответствующими весами, $y \in \{1, 0\}$, $\theta = [\theta, b]$

Модель дискриминации между наборами данных

$$d = G_d(G_f(\mathbf{x}, \theta_f, b_f), \theta_d, b_d),$$

$$G_d(G_f(\mathbf{x}), \theta_d, b_d) = \text{softmax}(\theta_d^T G_f(\mathbf{x}) + b_d),$$

где $G_d(\cdot, \theta_d)$ – дискриминатор, $d \in \{1, 0\}$, $d = 1$, если $\mathbf{x} \in S_1$, $d = 0$, если $\mathbf{x} \in S_2$.

Функция ошибки

$$\mathcal{L}(\theta_f, \theta_y, \theta_d) = \mathcal{L}_y(\theta_f, \theta_y) + \lambda \mathcal{L}_d(\theta_f, \theta_d)$$

$$\mathcal{L}_y(\theta_f, \theta_y) = \frac{1}{m} \sum_{i=1}^m \left[y_i \log \frac{1}{G_y(G_f(\mathbf{x}_i, \theta_f), \theta_y)} + (1-y_i) \log \frac{1}{1 - G_y(G_f(\mathbf{x}_i, \theta_f), \theta_y)} \right]$$

$$\mathcal{L}_d(\theta_f, \theta_y) = \frac{1}{2m} \sum_{i=1}^{2m} \left[d_i \log \frac{1}{G_d(G_f(\mathbf{x}_i, \theta_f), \theta_d)} + (1-d_i) \log \frac{1}{1 - G_d(G_f(\mathbf{x}_i, \theta_f), \theta_d)} \right]$$

где d_i есть метка номера набора данных для события (x_i, y_i) ,
 $d_i = 0$, если $\mathbf{x}_i \in S_1$ и $d_i = 1$ если $\mathbf{x}_i \in S_2$.

- $\mathcal{L}_y(\theta_f, \theta_y)$ ошибка классификации.
- $\mathcal{L}_d(\theta_f, \theta_y)$ ошибка дискриминации между двумя наборами данных.

Оптимизация составной функции ошибки

Функция ошибки

$$\mathcal{L}(\theta_f, \theta_y, \theta_d) = \mathcal{L}_y(\theta_f, \theta_y) + \lambda \mathcal{L}_d(\theta_f, \theta_d)$$

$\lambda > 0$ коэффициент, отвечающий за вклад дискриминатора

Подсчет градиента

$$\frac{\partial \mathcal{L}(\theta_f, \theta_y, \theta_d)}{\partial \theta_y} = \frac{\partial \mathcal{L}_y(\theta_f, \theta_y)}{\partial \theta_y}$$

$$\frac{\partial \mathcal{L}(\theta_f, \theta_y, \theta_d)}{\partial \theta_d} = \lambda \frac{\partial \mathcal{L}_d(\theta_f, \theta_d)}{\partial \theta_d}$$

$$\frac{\partial \mathcal{L}(\theta_f, \theta_y, \theta_d)}{\partial \theta_f} := \frac{\partial \mathcal{L}_y(\theta_f, \theta_y)}{\partial \theta_f} - \lambda \frac{\partial \mathcal{L}_d(\theta_f, \theta_d)}{\partial \theta_f}$$

Подсчет градиента по переменным θ_f видоизменяется, перед вторым слагаемым ставится минус с целью сблизить распределение признаков для двух распределений в скрытом пространстве.

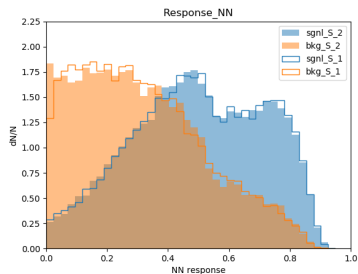
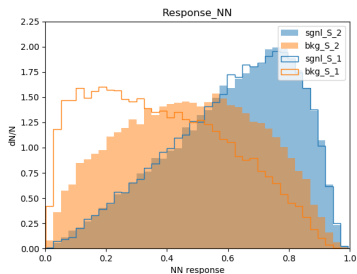
Для оценки качества модели классификации подсчитывается *значимость* Z_A . Учитывается дисбаланс классов на реальных данных (5% сигнал). Рассматриваются события со скором выше 0.6.

$$Z_A = \frac{s}{\sqrt{b_2 + \sigma_b^2}}$$

$s = 5 \cdot 10^4 \frac{s^{cut}}{s^{all}} \cdot 5\%$; s^{all} - число событий-сигналов из S_2 , s^{cut} - число событий-сигналов из S_2 со скором выше 0.6.

$b_{2(1)} = 5 \cdot 10^4 \frac{b_{2(1)}^{cut}}{b_{2(1)}^{all}} \cdot 95\%$; $\sigma_b = b_2 - b_1$; $b_{2(1)}^{all}$ - число шумовых событий, $b_{2(1)}^{cut}$ - число фоновых событий со скором выше 0.6.

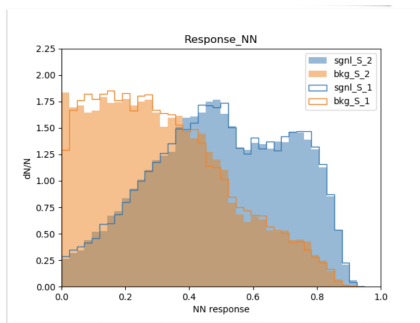
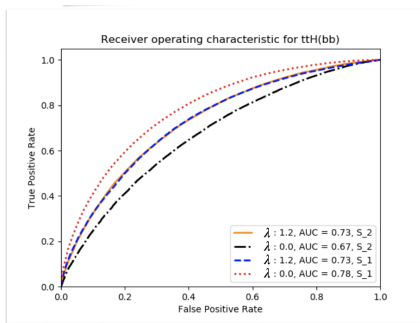
Гистограммы отклика



Разность между NN распределениями отклика на фон из разных наборов значительно уменьшилась. Число нейронов в слоях генератора признаков: 45, 30, 25; λ : 0 (слева), 1.2 (справа).

Распределения гистограммы отклика на обучающем и на валидационном наборе данных с другим распределением сблизились, причем зачимость выросла с 0.4 до 4.4.

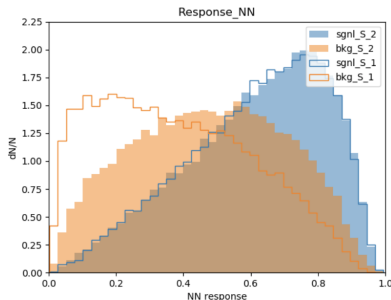
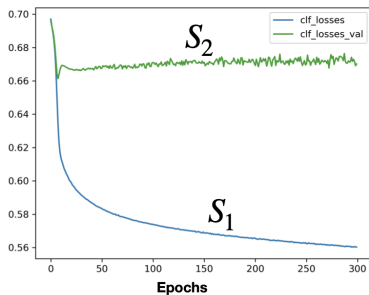
ROC-кривые



ROC-кривая и распределение отклика. Найден компромисс между точностью и обобщающей способностью модели в терминах площади под ROC-кривой.

Оптимизация функции ошибки в случае без дискриминатора

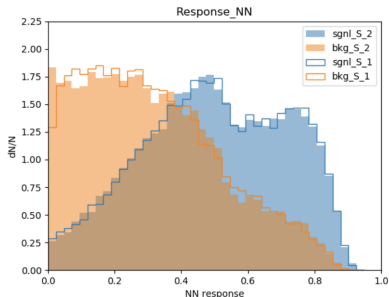
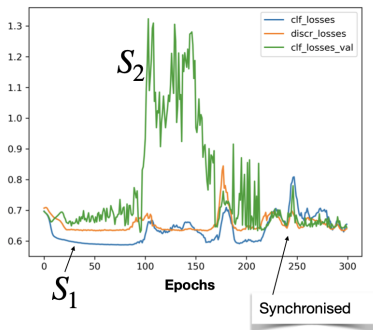
• $\lambda = 0$ Training plot



Оптимизация функции ошибки при $\lambda = 0$, то есть без дискриминатора.

Оптимизация функции ошибки с дискриминатором

- $\lambda = 1.20$ Training plot



Оптимизация функции ошибки при $\lambda = 1.2$. Заметно улучшение качества классификации при появлении состязательного поведения.

Исследовано применение подхода доменной состязательной адаптации в поиске бозона Хиггса в детекторе *ATLAS* на данных из Монте Карло генератора. Показано:

- данный подход уменьшает смещение при обучении на данных из генератора,
- показывает адекватное качество классификации.

Дальнейшие направления исследования:

- исследовать возможность применение к реальным данным,
- обобщить подход на случай нескольких Монте Карло генераторов.