

Машинное обучение при реконструкции сигналов с электромагнитного калориметра

Семенов А.В.

Август 27, 2026

Постановка задачи

Имеется процесс: $Z_T \xrightarrow{g(Z_T)} X_D \xrightarrow{g^{-1}(X_D)} Y = f(X_D) \approx Z_T$

Z_T — параметры частицы (энергия, углы) или истинные значения (Truth)

X_D — отклик детектора (Detector) (энерговыделение, вершина)

Y — реконструированные параметры, вычисляемые с помощью функции $f(X_D)$

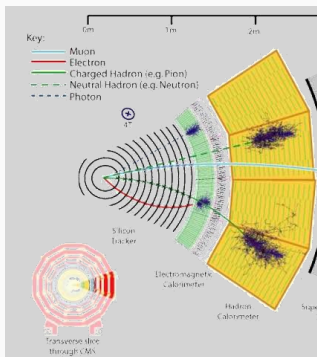
Классический подход: численное инвертирование функции отклика $g(Z_T) \rightarrow f(X_D) = g^{-1}(X_D)$,

где $g(Z_T) = \mathbb{E}_{train}[X_D | Z_T = z_T] = \int x \cdot p_{train}(x | z_T) dx$.

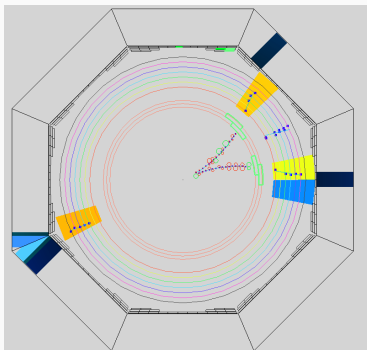
Возможно в случае $\dim X_D \leq 2$

Больше параметров $\rightarrow$ больше информации $\rightarrow$ лучше разрешение
Многопараметрическая функция $\rightarrow$ машинное обучение

Моделирование детектора



События CMS



Событие КМД-3

Ключевое предположение: $p_{sim}(x_D|z_T) = p_{exp}(x_D|z_T)$ — универсальность отклика детектора.

Используем то, что хорошо моделируется.

Geant4 $\rightarrow$ 1 млн фотонов

Топология сети:

$5 \rightarrow 5 \rightarrow 10 \rightarrow 10 \rightarrow 5 \rightarrow 1$

Вход (нормируем):

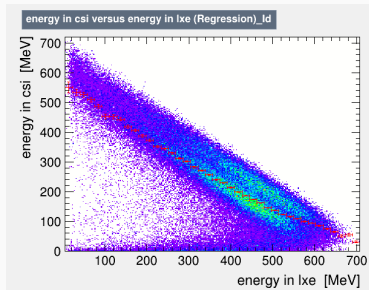
- E_{LXe} — энергия в LXe
- E_{CsI} — энергия в CsI
- ρ — радиус конверсии
- θ — угол вылета
- ϕ — угол вылета

Выход (нормируем):

- E_γ — энергия фотона

Функции активации: ELU

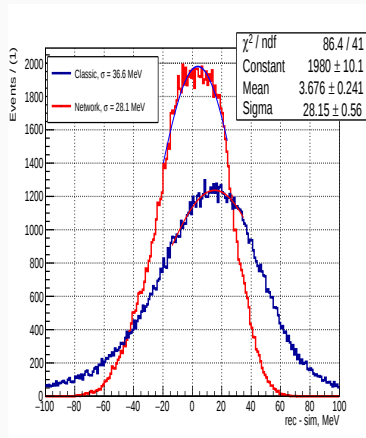
$$\mathcal{L} = \sum_i (z_i - f(x_i))^2 \rightarrow \min$$



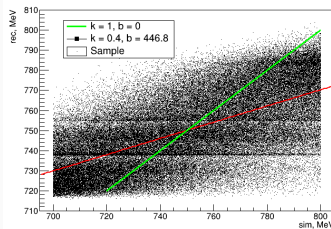
Корреляция CsI и LXe

Тренируем на фотонах с энергией $[700, 800] \text{ MeV}$

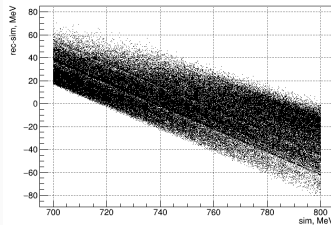
Не используем энергию по слоям и т.д.:
плохо моделируется.



Сравнение отклонений



rec от sim



Отклонение от sim

Причина

Получившееся отклонение (28MeV) заведомо меньше ожидаемого значения (36MeV), нейронная сеть решила задачу поворотом (минимизация проекции $z - f(x)$).

$f_{MSE} = \underset{g}{\operatorname{argmin}} [(g(X_D) - Z_T)^2]$ и если сеть достаточно хороша $\implies$

$$f_{MSE}(X_D) = \mathbb{E}_{\text{train}} [Z_T | X_D] = \int z \cdot p_{\text{train}}(z | X_D) dz = \int z \cdot p(X_D | z) \frac{p(z)}{p(X_D)} dz$$

Функция явно зависит от приорного распределения $p(z)$, а также $p_{\text{train}}(z) \neq p_{\text{test}}(z)$, что создаёт явное смещение.

$$b(z_T) = \mathbb{E}_{\text{test}} [f(X_D - z_T | Z_T = z_T)]$$

$$b(z_T) + z_T = \int f(x_D) \cdot p_{\text{test}}(x_D | z_T) dx_D = \int z \cdot p_{\text{train}}(z | x_D) p_{\text{test}}(x_D | z_T) dz dx_D$$

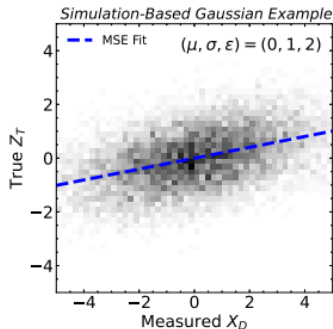
Смещение есть даже при $p_{\text{train}} = p_{\text{test}}$

Пример №1

$$Z_T \sim \mathcal{N}(\mu, \sigma^2), \text{ а } X_D | [Z_T = z_T] \sim \mathcal{N}(z_T, \epsilon^2) \implies \\ X_D \sim \mathcal{N}(\mu, \sigma^2 + \epsilon^2)$$

$$f_{MSE}(x_D) = \frac{\epsilon^2 \mu + \sigma^2 x_D}{\epsilon^2 + \sigma^2}$$

$$\Sigma_{MSE}(x_D) = \frac{\epsilon \sigma}{\sqrt{\epsilon^2 + \sigma^2}}$$

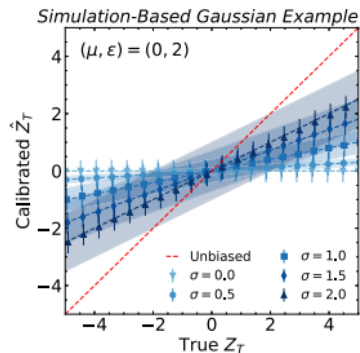
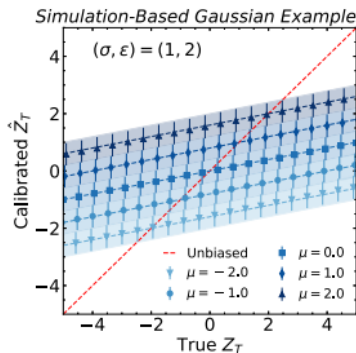


Данные и результат обучения

Ожидаем $f(x_D) = x_D$ и $\Sigma(x_D) = \epsilon$

Пример №1

$$b(z_T) = \frac{\epsilon^2}{\epsilon^2 + \sigma^2}(\mu - z_T)$$



Зависимость смещения от параметров моделирования

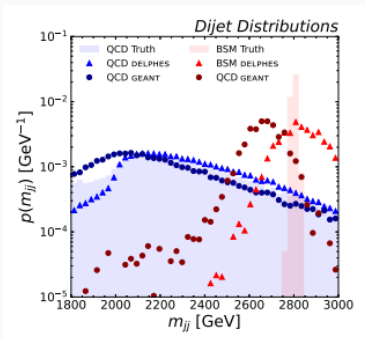
Как спрятать смещение: расширить область моделирования (σ)

Пример №2

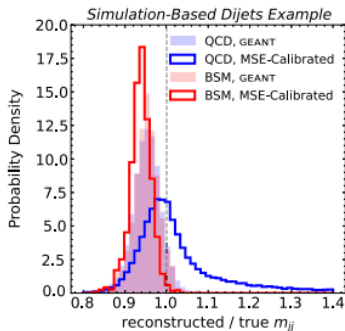
Рассмотрим пример, когда $p_{sim} \neq p_{test}$

Обучаем реконструкции на полном спектре (синии точки)

Тестируем на узкой области



Моделирование



Тестирование

Современные решения используют следующий подход:

$$f_{MLC}(x) = \underset{z}{argmax} p(x|z)$$

Для каких фотонов заданный отклик наиболее вероятен

$$\mathcal{L} = -\mathbb{I}(X; Z) = - \int p(x, z) \cdot \log \frac{p(x, z)}{p(x)p(z)} dx dz$$

Максимизация взаимной информации

Взаимная информация равняется 0, если величины независимые

$$\mathcal{L}_{DVR}[T] = - \left(\mathbb{E}_{P_{XZ}}[T] - \log \left(\mathbb{E}_{P_X \otimes P_Z}[e^T] \right) \right)$$

$$T(x, z) = \log \frac{p(x|z)}{p(x)} + c$$

$$f_{MLC}(x) = \underset{z}{\operatorname{argmax}} T(x, z)$$

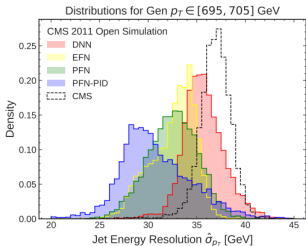
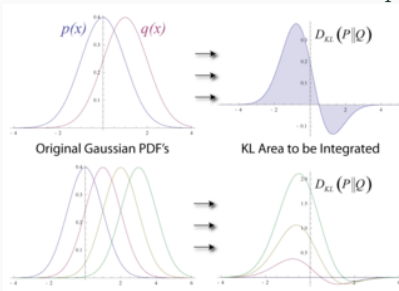


FIG. 1. Learned JER distribution for the four models, compared to the CMS 2011 baseline. The dataset is the same as in Table I. On average, the PFN-PID exhibits 15% better resolution (i.e. smaller values) than the CMS default.

Критика: максимум взаимной информации необязательно сопоставляет $X \longleftrightarrow Z$, максимум возможен при $x^3 = z$.

$$D_{KL}(P||Q) = \int_Y p(y) \log \frac{p(y)}{q(y)} dy$$

“Расстояние” между распределениями. Всегда положительная, несимметричная мера.



Q — аппроксимирующая функция, P — имеющееся распределение

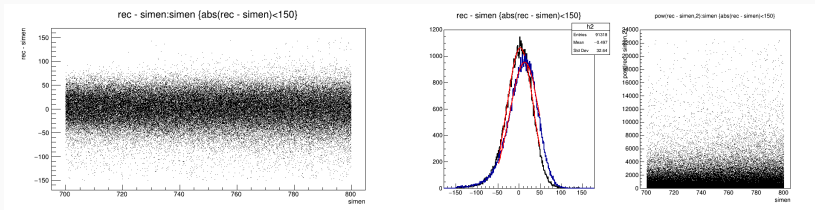
$$D_{KL}(P||Q) = \int_Y p(y) \log \frac{p(y)}{q(y)} dy$$

$$\mathcal{L} = \int D_{KL}(P_{\text{apriori}}(X|Z)||P_{\text{est}}(X|Z)) dZ$$

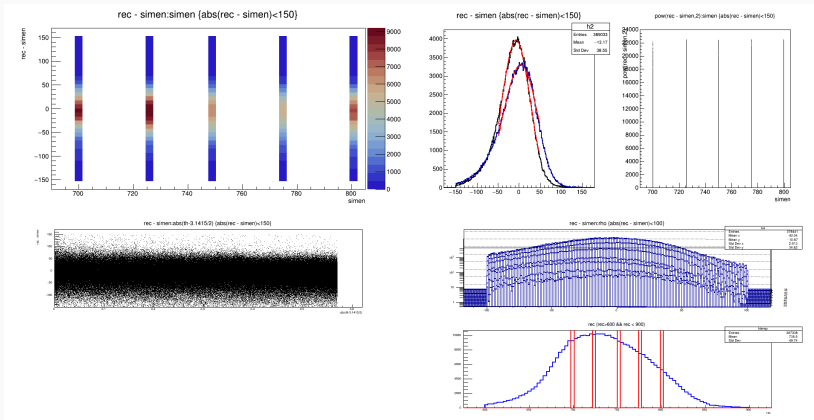
P_{est} оценивается из KDE: $\sum_i K(\frac{y-f(x_i)}{h})$

P_{apriori} — нормальное распределение: $\mathcal{N}(z, \sigma^2(x))$

Улучшение: $\sigma : 34.7 \rightarrow 30.2 \text{ MeV}, \mu : 12.9 \rightarrow 2.6 \text{ MeV}$



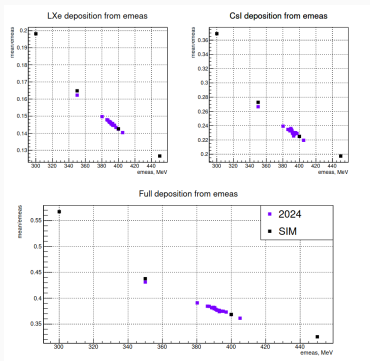
Отобрал события $e^+e^- \rightarrow \gamma\gamma$ Улучшение:
 $\sigma : 40.9 \rightarrow 32.6 \text{ MeV}, \mu : 3.4 \rightarrow -7.2 \text{ MeV}$



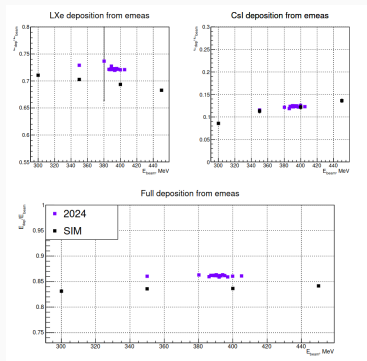
- Всегда проверять, что делает NN!
- Если используется MSE, скорее всего всё неправильно
- Современные подходы: NN восстанавливает распределение отклика детектора
- Аналогично можно сказать про классификацию

Статьи:

- <https://arxiv.org/abs/2205.05084>
- <https://arxiv.org/abs/2205.03413>
- <https://arxiv.org/abs/1801.04062>
- Внутри работ большое кол-во ссылок на другие работы, которые стоит посмотреть



MIP



Bhabha

Энерговыделение в цилиндрическом калориметре КМД-3 2025-2026 год