

Содержание

1. Введение	4
2. Детектор КМД-3 и LHe калориметр	5
3. Формат входных данных с калориметра	7
4. Библиотека TMVA и метод BDT	7
5. Алгоритм разделения кластеров	9
6. Результаты	10
7. Заключение	10
Список литературы	10

1. Введение

В Институте ядерной физики им. Г. И. Будкера СО РАН проводятся эксперименты на электрон-позитронном коллайдере ВЭПП-2000. Коллайдер имеет два места встречи пучков, в которых установлены детекторы КМД-3 (криогенный магнитный детектор) и СНД (сферический нейтральный детектор). В программу проекта ВЭПП-2000 включено измерение сечений процессов, наблюдаемых при столкновениях электронов и позитронов в диапазоне энергий до 2 ГэВ. В число изучаемых процессов входят процессы с фотонами в конечном состоянии [1].

Для регистрации фотонов в детекторе КМД-3 используется жидкоксеноновый (LXe) калориметр. Фотон, попадая в калориметр, порождает электромагнитный ливень из вторичных электронов, позитронов и фотонов. Частицы ливня регистрируются сигнальными полосками, расположенными на слоях калориметра. Сработавшие сигнальные полоски группируются в кластеры.

При столкновениях электронов e^- и позитронов e^+ большое сечение рождения имеют процессы, содержащие нейтральный пион π^0 в конечном состоянии. Одним из таких процессов является $e^-e^+ \rightarrow \pi^0\gamma$. Нейтральный пион с вероятностью 99% распадается на два фотона γ . Минимальный угол между фотонами при более высокой энергии пи-мезона становится меньше, и при значении $E_{\pi^0} = 1$ ГэВ этот угол составляет 0,27 радиана. Электромагнитные ливни таких фотонов в калориметре могут перекрываться, и появляется вероятность принять два слившихся кластера, образованных двумя фотонами, за одиночный кластер от одного фотона. Чтобы увеличить статистику событий, повысить точность измерения сечения и других параметров процесса, необходимо разделять такие кластеры [2].

Таким образом, требуется разработать алгоритм различения однофотонных и двухфотонных кластеров в LXe калориметре детектора КМД-3. Для решения этой задачи выполнены следующие шаги:

- описание детектора КМД-3 и LXe калориметра,
- подготовка данных для обработки,
- обзор библиотеки TMVA 4 и метода BDT,
- анализ задачи разделения кластеров,
- реализация алгоритма разделения кластеров с помощью метода BDT.

Результаты данной работы планируется сравнить с результатами, полученными в работах [2, 3].

2. Детектор КМД-3 и LXe калориметр

На рис. 1 представлена схема криогенного магнитного детектора КМД-3.

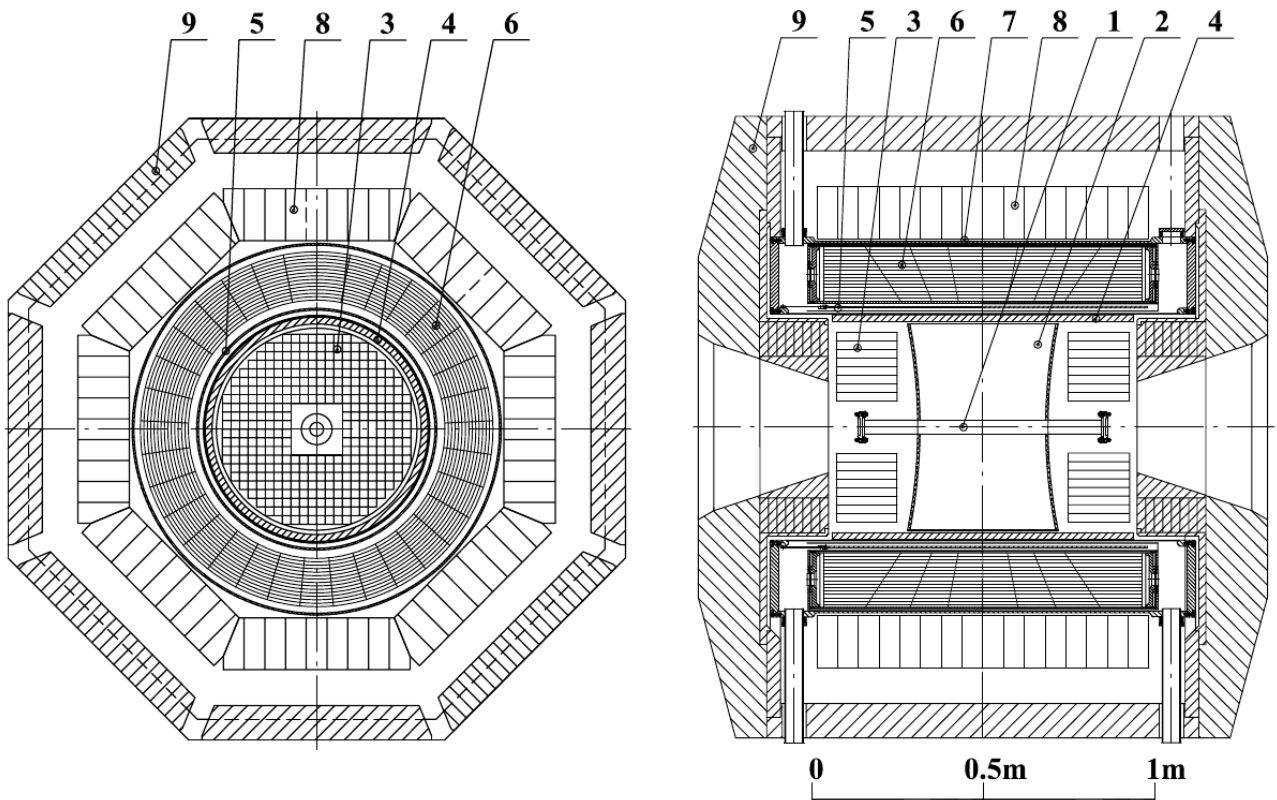


Рис. 1. Схема детектора КМД-3: 1 – вакуумная камера, 2 – дрейфовая камера, 3 – торцевой BGO калориметр, 4 – Z-камера, 5 – сверхпроводящий соленоид, 6 – цилиндрический жидкоксеноновый калориметр, 7 – времяпролётная система, 8 – цилиндрический CsI калориметр, 9 – ярмо магнита [4].

Цилиндрический электромагнитный калориметр детектора регистрирует фотоны и электроны, которые вылетают под большими углами к оси пучков (от 39° до 141°), и охватывает телесный угол $0,8 \cdot 4\pi$. Он состоит из двух

соосных подсистем: жидкоксенонового LXe калориметра и кристаллического CsI калориметра.

Жидкоксеноновый калориметр образован семью соосными катодами и восемью анодами, образующими систему ионизационных камер. Структура электродов LXe калориметра представлена на рис. 2.

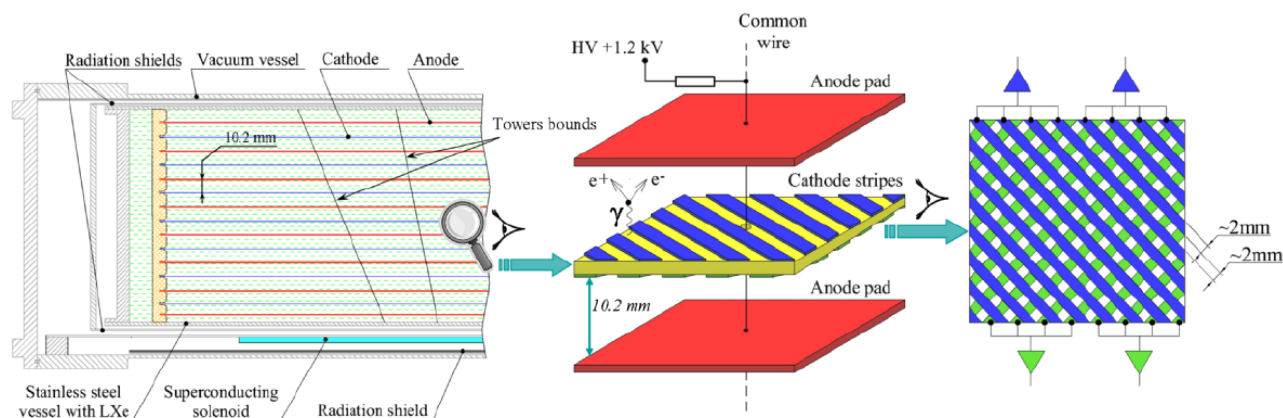


Рис. 2. Структура электродов жидкоксенонового калориметра [4].

Поверхность каждого анода разделена на 264 прямоугольные площадки – восемь вдоль оси пучков Z , тридцать три в перпендикулярной ей плоскости R, φ . Площадки с одинаковым положением по Z и φ соединены проводником в «башню». Такая «башня» ориентирована в место встречи пучков. Аноды используются для измерения выделенной энергии.

Каждый катод разбит на 2112 полосок. Полоски объединены в группы по четыре, и с каждой такой группы сигнал снимается по одному каналу. На противоположных сторонах катода полоски ориентированы взаимно перпендикулярно и находятся под углом 45° к оси пучков Z . Такая структура катода используется для восстановления координат кластера [4].

Регистрация фотонов производится путём сбора информации о частицах электромагнитного ливня, порождённого при взаимодействии фотона с материалом калориметра. LXe калориметр позволяет измерять координаты с погрешностью от 0,7 мм для первого слоя до 1,4 мм для шестого слоя [5, 6]. Точность измерения энергии составляет примерно 8,5% для нейтральных пи-мезонов [7].

3. Формат входных данных с калориметра

4. Библиотека TMVA и метод BDT

Библиотека TMVA (The Toolkit for Multivariate Analysis) предоставляет среду для параллельного вычисления и применения методов многомерной классификации и, начиная с версии TMVA 4, многомерной регрессии. В данной библиотеке представлены только алгоритмы машинного обучения с учителем. Они используют тренировочные события с известным результатом, чтобы определить отображение, которое описывает либо критерии классификации, либо регрессионное приближение для функции результата. Среда TMVA разработана в основном для нужд физики высоких энергий, но может применяться и в других областях [8].

В данной работе используется метод BDT (boosted decision trees), основанный на нахождении множества деревьев решений с присвоенными им весами. Дерево решений – это классификатор, построенный в виде двоичного дерева. Пример такого дерева изображён на рис. 3.

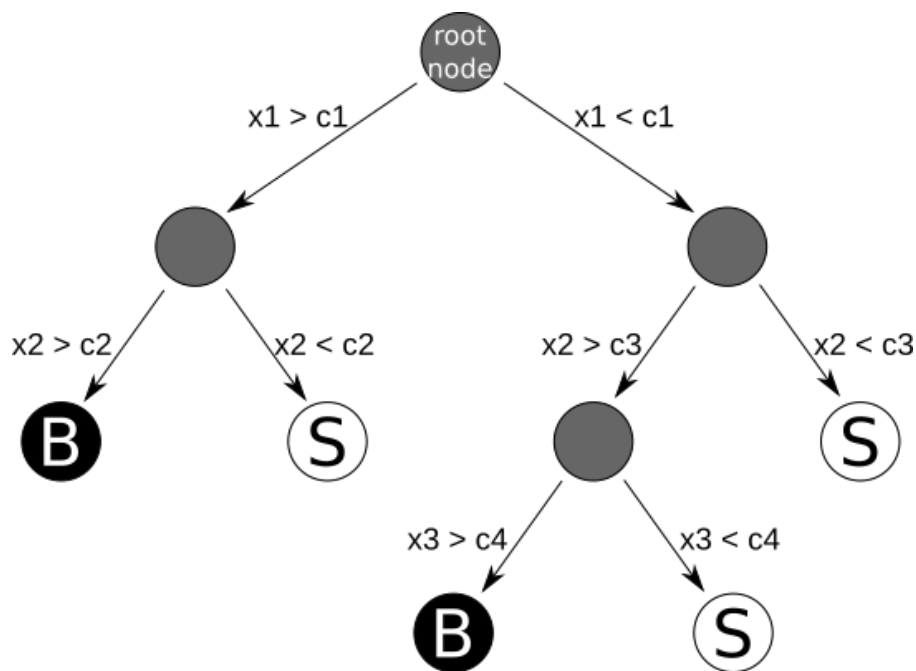


Рис. 3. Пример дерева решений. Решения принимаются на основе переменных x_1 , x_2 , x_3 , критериями являются величины c_1 , c_2 , c_3 , c_4 .

Задача алгоритма, использующего дерево решений, заключается в том,

чтобы пройти это дерево, начиная с корневого узла, и закончить на одном из листьев дерева. В каждом узле находится критерий выбора дочернего узла по значению одной из переменных. В результате алгоритм относит событие к категории сигнальных или фоновых.

Тренировка дерева производится поэтапно, начиная с корневого узла. Выбирается критерий, наиболее эффективно делящий события на фоновые и сигнальные, то есть дающий наибольшую долю правильно классифицированных событий в обеих категориях. Такая же процедура выбора критерия выполняется на каждом из двух образовавшихся подмножеств событий и так далее до определённой глубины дерева. Таким образом, фазовое пространство оказывается поделено на сегменты, помеченные как сигнал или фон в зависимости от того, событий какого типа больше всего попало в данный сегмент при тренировке алгоритма.

Достоинства дерева решений – простота описания и интерпретации, а также возможность разбивать фазовое пространство на большое число гиперкубов, обозначаемых как фон или сигнал – в отличие от анализа на основе совокупности пороговых значений, которая выделяет лишь один гиперкуб. Недостатком дерева решений является его чувствительность к статистическим флуктуациям в тренировочном наборе событий. Например, две переменные, одинаково влияющие на классификацию событий, под действием флуктуаций могут идти в разном порядке следования сравнений, что изменит и всю дальнейшую структуру дерева, в итоге воздействуя на результат классификации.

Для уменьшения флуктуаций, метод BDT использует бустинг деревьев. Для этого данные делятся на некоторое количество ансамблей N , и на каждом ансамбле тренируется по одному дереву решений. В итоге получается N деревьев, каждое из которых имеет свой уровень разделения и даёт свой отклик на одно и то же событие. Деревьям присваиваются веса в зависимости от

качества разделения. Итоговая величина r отклика на событие вычисляется как средневзвешенное по всем деревьям:

$$r = \sum_i \omega_i r_i,$$

где ω_i – вес i -го дерева, r_i – отклик i -го дерева [8].

5. Алгоритм разделения кластеров

Задача алгоритма – имея координаты совокупности точек и связанные с ними значения энергии, классифицировать событие как сигнальное или фоновое. Сигнальным событием считается появление кластера, сгенерированного только одним фотоном, а фоновое событие – это регистрация слившихся кластеров.

Для решения этой задачи в среде C++/ROOT с использованием библиотеки TMVA 4 разработаны следующие процедуры:

1. генерация набора данных для обучения и проверки алгоритма,
2. реализация алгоритма при помощи метода BDT,
3. проверка работы алгоритма на некотором наборе данных.

Генератор тренировочного набора данных на выходе даёт заданное количество событий, имеющих три параметра: массу частицы, номер катодного слоя, зарегистрировавшего частицу, и амплитуду сигнала. Каждому событию присваивается тип – сигнальное или фоновое событие (рис. 4).

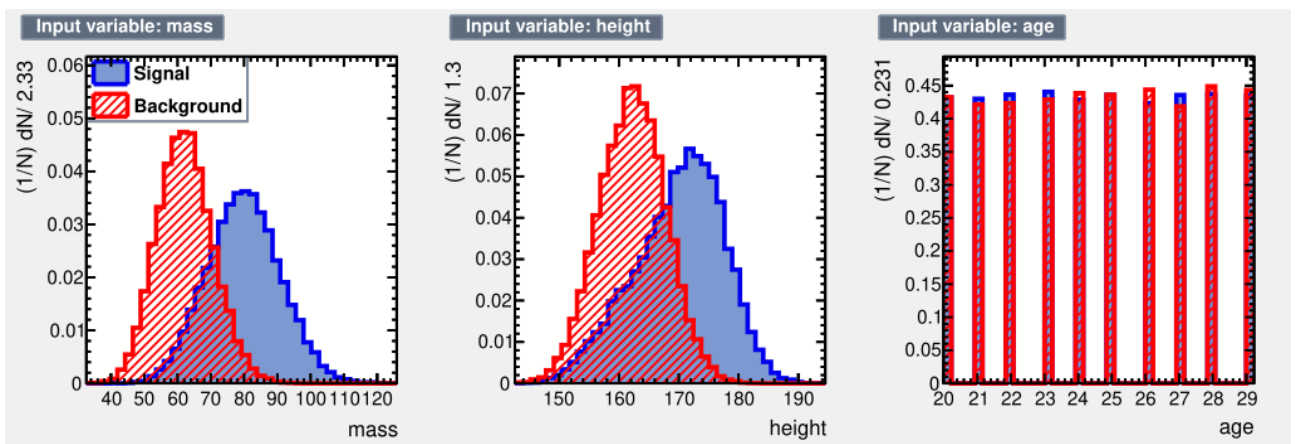


Рис. 4. Распределения входных данных. Сгенерировано 100000 событий. Переменная *mass* – масса частицы, *height* – амплитуда сигнала, *age* – номер слоя.

6. Результаты

7. Заключение

Список литературы

- [1] Yu. M. Shatunov et al. Project of a new electron positron collider VEPP-2000. *Conf. Proc.*, C0006262:439–441, 2000.
- [2] Лакомов А. Е. Разделение близких фотонов в калориметре КМД-3. Выпускная квалификационная работа, НГУ, Новосибирск, 2019.
- [3] Рабусов А. В. Изучение процесса аннигиляции $e^+e^- \rightarrow \eta \gamma, \eta \pi^+\pi^-\pi^0$ с детектором КМД-3 на коллайдере ВЭПП-2000. Магистерская диссертация, НГУ, Новосибирск, 2016.
- [4] Шебалин В. Е. *Реконструкция фотонов и энергетическая калибровка цилиндрического калориметра детектора КМД-3*. Диссертация на соискание учёной степени кандидата физико-математических наук, ИЯФ СО РАН, Новосибирск, 2016.
- [5] Denis Epifanov. Electromagnetic calorimeters of the CMD-3 detector. *Journal of Physics: Conference Series*, 293:012009, April 2011.
- [6] A. V. Anisyonkov et al. Liquid xenon calorimeter for a CMD-3 detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 598:266–267, January 2009.
- [7] A. V. Anisenkov et al. Status of the Liquid Xenon calorimeter of the CMD-3 detector. *Journal of Instrumentation*, 9(08):C08024–C08024, August 2014.
- [8] A. Hoecker et al. *TMVA 4. Toolkit for Multivariate Data Analysis with ROOT. Users Guide*. Regents of CERN et al., 2005-2009.