

Ереванский Государственный Университет

физический факультет

кафедра ядерной физики

Согомонян Арсен

Дипломная работа

тема: Обработка данных атмосферных черенковских телескопов

руководитель: Чилингарян А.А.

Ереван 2002

Введение

Первичное космическое излучение достигающее земли состоит в основном из протонов и ядер среди которых есть малая доля гамма-излучения. Так как гамма-кванты не имеют электрического заряда то они достигают земли не отклоняясь в магнитных полях, и являются основным источником информации об объектах которые их излучают. Поэтому изучение гамма-излучения представляет большой интерес.

Источниками гамма-квантов являются активные ядра галактик, остатки суперновых звезд, нейтронные звезды, радиогалактику, пульсары и др. Гамма-излучение этих объектов возникает в основном за счёт обратного комптоновского рассеяния высокоэнергетичных электронов на фотонах или при распаде нейтральных пи-мезонов образовавшихся при взаимодействии протонов с газом вблизи звезд. Исследования показали что в каждой галактике имеется активный центр в котором разрастается черная дыра, которая при вращении излучает гамма-кванты в двух противоположных направлениях параллельных оси вращения (рис. 1).

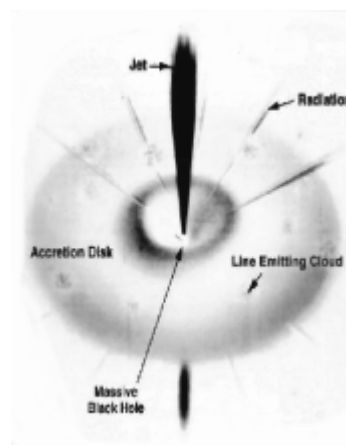


рис.1 излучение черной дыры

Исследование потоков гамма-квантов в области промежуточных энергий 10^{11} - 10^{13} эв является недоступным ни для спутниковых экспериментов из-за слишком слабых потоков ни для крупных наземных сцинтилляционных установок. ввиду затухания широких атмосферных ливней инициированных частицами этих энергий, т.к. Гамма-кванты поглощаются в атмосфере земли и очень малая их часть достигает поверхности земли.

Поглощаясь в атмосфере гамма-кванты генерируют широкие атмосферные ливни. Вторичные частицы ливня имея большую энергию (и скорость) генерируют черенковское излучение. Черенковское излучение

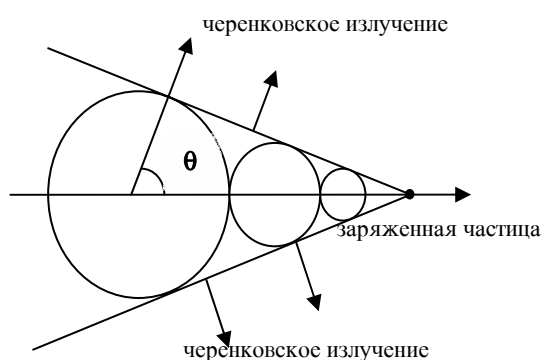


рис.2 $\cos \theta = \frac{1}{\beta n}$, где n – показатель преломления среды.

возникает при движении заряженных частиц в среде со скоростью превышающую скорость света в данной среде. Черенковский свет излучается под определенным характеристическим углом по направлению к движению частицы (см. рис.2)

Черенковский свет от ШАЛ может быть зарегистрирован с помощью черенковских атмосферных телескопов. Черенковские телескопы представляют собой детекторы установленные на поверхности земли предназначенные для регистрации черенковского излучения возникающего при развитии широких атмосферных ливней. Они состоят из оптического рефлектора который направляет собранный свет на ФЭУ расположенный в фокусе.

Телескоп MAGIC

Одним из черенковских атмосферных телескопов является телескоп MAGIC расположенный на Канарских островах (рис.3). Зеркальная поверхность телескопа имеет площадь 234 кв.м. и состоит из квадратных элементов размером 50x50 см.

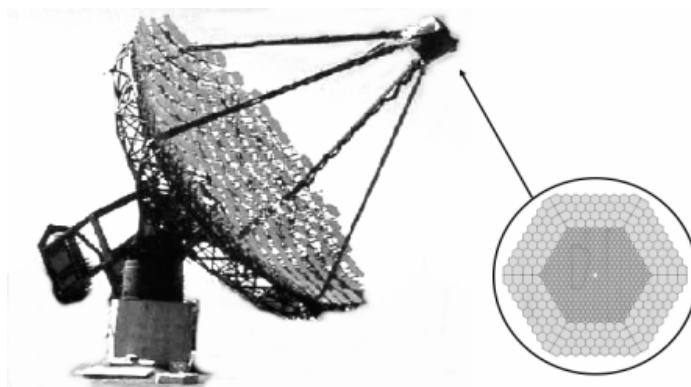


рис.3 Телескоп MAGIC

Зеркала направляют свет на камеру имеющую диаметр 1.5м.

В центре камеры расположены 397 фотоумножителей имеющих диаметр ~2.5см а по периметру 180 фотоумножителей имеющих диаметр ~3,8см. Порог регистрации гамма-квантов фотоумножителем составляет 30 GeV.

Основным препятствием к применению черенковской техники является очень большой фон от ШАЛ инициированных протонами и ядрами космического излучения (адронных ливней). Поэтому большое значение имеют работы по подавлению фона от адронных ливней.

Количество мюонов в адронных ливнях примерно в 30 раз больше чем в ливнях образованных гамма-частицами. Таким образом можно использовать для разделения

ливней инициированных гамма-частицами и адронами детекторы мюонов. Но так как мюоны покрывают большие поверхности этот метод является неэффективным. Другое различие адронных и гамма-ливней состоит в том что адронные ливни, обычно шире и их развитие и затухание происходит на меньшей высоте по сравнению с гамма-ливнями. Ливни инициированные гамма-квантами излучаемыми определенным объектом имеют направление совпадающее с направлением к данному объекту.

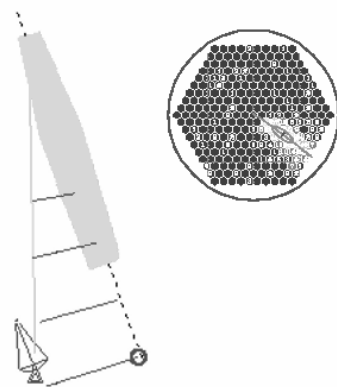


рис.4 ливень и его изображение

При ориентации телескопа по направлению ливня изображение ливня на камере телескопа имеет вид эллипса(рис.4). Положение, ориентация и размеры эллипса характеризуются рядом параметров.(рис.5).

Классификация ливней по этим параметрам дает хорошие результаты. Одним из методов классификации ливней с использованием упомянутых выше параметров является использование нейронных сетей.

Нейронные сети

В последние десятилетия в мире бурно развивается новая прикладная область математики, специализирующаяся на искусственных нейронных сетях (НС). Актуальность исследований в этом направлении подтверждается массой различных применений НС. Это автоматизация процессов распознавания образов, адаптивное управление, аппроксимация функционалов, прогнозирование, создание экспертных систем, организация ассоциативной памяти и многие другие приложения. С помощью НС можно, например, предсказывать показатели биржевого рынка, выполнять распознавание оптических или звуковых сигналов, создавать самообучающиеся системы. Широкий круг задач, решаемый НС, не позволяет в настоящее время создавать универсальные, мощные сети, вынуждая разрабатывать специализированные НС, функционирующие по различным алгоритмам.

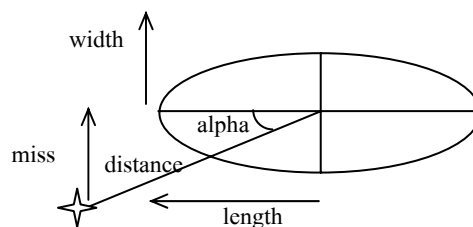


рис.5 параметры изображения

Несмотря на существенные различия, отдельные типы НС обладают несколькими общими чертами.

Основу каждой НС составляют относительно простые, в большинстве случаев

– однотипные, элементы (ячейки), имитирующие работу нейронов мозга. Далее под нейроном будет подразумеваться искусственный нейрон, то есть ячейка НС. Каждый нейрон характеризуется своим текущим состоянием по аналогии с нервными клетками головного мозга, которые могут быть возбуждены или заторможены. Он обладает

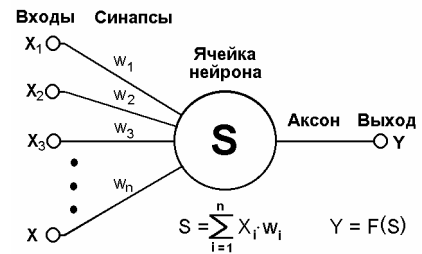


Рис.6 Искусственный нейрон

группой синапсов – однонаправленных входных связей, соединенных с выходами других нейронов, а также имеет аксон – выходную связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов. Общий вид нейрона приведен на рисунке 1. Каждый синапс характеризуется величиной синаптической связи или ее весом w_i , который по физическому смыслу эквивалентен электрической проводимости.

Текущее состояние нейрона определяется, как взвешенная сумма его входов:

$$s = \sum_{i=1}^n x_i \cdot w_i \quad (1)$$

Выход нейрона есть функция его состояния:

$$y = f(s) \quad (2)$$

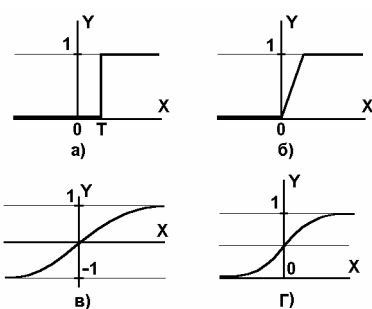


Рис.2 а) функция единичного скачка; б) линейный порог (гистерезис); в) сигмоид – гиперболический тангенс; г) сигмоид – формула (3)

Нелинейная функция f называется активационной и может иметь различный вид, как показано на рисунке 2. Одной из наиболее распространенных является нелинейная функция с насыщением, так называемая логистическая функция или сигмоид (т.е. функция S-образного вида):

$$f(x) = \frac{1}{1 + e^{-\alpha x}} \quad (3)$$

При уменьшении α сигмоид становится более пологим, в пределе при $\alpha=0$ вырождаясь в горизонтальную линию на уровне 0.5, при увеличении α сигмоид приближается по внешнему виду к функции единичного скачка с порогом T в точке $x=0$. Из

выражения для сигмоида очевидно, что выходное значение нейрона лежит в диапазоне $[0,1]$. Одно из ценных свойств сигмоидной функции – простое выражение для ее производной, применение которого будет рассмотрено в дальнейшем.

$$f'(x) = \alpha \cdot f(x) \cdot (1 - f(x)) \quad (4)$$

Следует отметить, что сигмоидная функция дифференцируема на всей оси абсцисс, что используется в некоторых алгоритмах обучения. Кроме того она обладает свойством усиливать слабые сигналы лучше, чем большие, и предотвращает насыщение от больших сигналов, так как они соответствуют областям аргументов, где сигмоид имеет пологий наклон.

Из общей чертой, присущей всем НС является принцип параллельной обработки сигналов, который достигается путем объединения большого числа нейронов в так называемые слои и соединения определенным образом нейронов различных слоев, а также, в некоторых конфигурациях, и нейронов одного слоя между собой, причем обработка взаимодействия всех нейронов ведется послойно.

В качестве примера простейшей НС рассмотрим трехнейронный перцептрон (рис.3), то есть такую сеть, нейроны которой имеют активационную функцию в виде единичного скачка*. На n входов поступают некие сигналы, проходящие по синапсам на 3 нейрона, образующие единственный слой этой НС и выдающие три выходных сигнала:

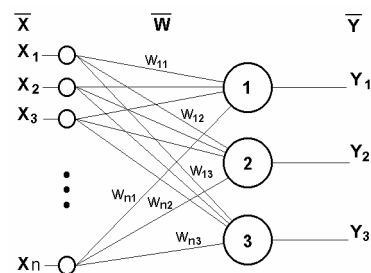


Рис.7 Однослойный перцептрон

$$y_j = f \left[\sum_{i=1}^n x_i \cdot w_{ij} \right], \quad j=1 \dots 3 \quad (5)$$

Очевидно, что все весовые коэффициенты синапсов одного слоя нейронов можно свести в матрицу \mathbf{W} , в которой каждый элемент w_{ij} задает величину i -ой синаптической связи j -ого нейрона. Таким образом, процесс, происходящий в НС, может быть записан в матричной форме:

$$\mathbf{Y} = \mathbf{F}(\mathbf{XW}) \quad (6)$$

где \mathbf{X} и \mathbf{Y} – соответственно входной и выходной сигнальные векторы, $\mathbf{F}(\mathbf{V})$ – активационная функция, применяемая поэлементно к компонентам вектора \mathbf{V} .

Теоретически число слоев и число нейронов в каждом слое может быть произвольным, однако фактически оно ограничено ресурсами компьютера или специализированной микросхемы, на которых обычно реализуется НС. Чем сложнее НС, тем масштабнее задачи, подвластные ей.

Выбор структуры НС осуществляется в соответствии с особенностями и сложностью задачи. Очевидно, что процесс функционирования НС, то есть сущность действий, которые она способна выполнять, зависит от величин синаптических связей, поэтому, задавшись определенной структурой НС, отвечающей какой-либо задаче, разработчик сети должен найти оптимальные значения всех переменных весовых коэффициентов (некоторые синаптические связи могут быть постоянными).

Этот этап называется обучением НС, и от того, насколько качественно он будет выполнен, зависит способность сети решать поставленные перед ней проблемы во время эксплуатации. На этапе обучения кроме параметра качества подбора весов важную роль играет время обучения. Как правило, эти два параметра связаны обратной зависимостью и их приходится выбирать на основе компромисса.

Обучение НС может вестись с учителем или без него. В первом случае сети предъявляются значения как входных, так и желательных выходных сигналов, и она по некоторому внутреннему алгоритму подстраивает веса своих синаптических связей. Во втором случае выходы НС формируются самостоятельно, а веса изменяются по алгоритму, учитывающему только входные и производные от них сигналы.

Существует великое множество различных алгоритмов обучения, которые однако делятся на два больших класса: детерминистские и стохастические. В первом из них подстройка весов представляет собой жесткую последовательность действий, во втором – она производится на основе действий, подчиняющихся некоторому случайному процессу.

Алгоритм обратного распространения

Среди различных структур нейронных сетей (НС) одной из наиболее известных является многослойная структура, в которой каждый нейрон произвольного слоя связан со всеми аксонами нейронов предыдущего слоя или, в

случае первого слоя, со всеми входами НС. Такие НС называются полносвязными. Когда в сети только один слой, алгоритм ее обучения с учителем довольно очевиден, так как правильные выходные состояния нейронов единственного слоя заведомо известны, и подстройка синаптических связей идет в направлении, минимизирующем ошибку на выходе сети. По этому принципу строится, например, алгоритм обучения однослойного перцептрона. В многослойных же сетях оптимальные выходные значения нейронов всех слоев, кроме последнего, как правило, не известны, и двух или более слойный перцептрон уже невозможно обучить, руководствуясь только величинами ошибок на выходах НС. Один из вариантов решения этой проблемы – разработка наборов выходных сигналов, соответствующих входным, для каждого слоя НС, что, конечно, является очень трудоемкой операцией и не всегда осуществимо. Вторым вариантом – динамическая подстройка весовых коэффициентов синапсов, в ходе которой выбираются, как правило, наиболее слабые связи и изменяются на малую величину в ту или иную сторону, а сохраняются только те изменения, которые повлекли уменьшение ошибки на выходе всей сети. Очевидно, что данный метод, несмотря на свою кажущуюся простоту, требует громоздких рутинных вычислений. И, наконец, третий, более приемлемый вариант – распространение сигналов ошибки от выходов НС к ее входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы. Этот алгоритм обучения НС получил название процедуры обратного распространения.

Согласно методу наименьших квадратов, минимизируемой целевой функцией ошибки НС является величина:

$$E(w) = \frac{1}{2} \sum_{j,p} (y_{j,p}^{(N)} - d_{j,p})^2 \quad (1)$$

где $y_{j,p}^{(N)}$ – реальное выходное состояние нейрона j выходного слоя N нейронной сети при подаче на ее входы p -го образа; $d_{j,p}$ – идеальное (желаемое) выходное состояние этого нейрона.

Суммирование ведется по всем нейронам выходного слоя и по всем обрабатываемым сетью образам. Минимизация ведется методом градиентного спуска, что означает подстройку весовых коэффициентов следующим образом:

$$\Delta w_{ij}^{(n)} = -\eta \cdot \frac{\partial E}{\partial w_{ij}} \quad (2)$$

Здесь w_{ij} – весовой коэффициент синаптической связи, соединяющей i -ый нейрон слоя $n-1$ с j -ым нейроном слоя n , η – коэффициент скорости обучения, $0 < \eta < 1$.

Как известно,

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} \cdot \frac{dy_j}{ds_j} \cdot \frac{\partial s_j}{\partial w_{ij}} \quad (3)$$

Здесь под y_j , как и раньше, подразумевается выход нейрона j , а под s_j – взвешенная сумма его входных сигналов, то есть аргумент активационной функции. Так как множитель dy_j/ds_j является производной этой функции по ее аргументу, из этого следует, что производная активационной функция должна быть определена на всей оси абсцисс. В связи с этим функция единичного скачка и прочие активационные функции с неоднородностями не подходят для рассматриваемых НС. В них применяются такие гладкие функции, как гиперболический тангенс или классический сигмоид с экспонентой. В случае гиперболического тангенса

$$\frac{dy}{ds} = 1 - s^2 \quad (4)$$

Третий множитель $\partial s_j / \partial w_{ij}$, очевидно, равен выходу нейрона предыдущего слоя $y_i^{(n-1)}$.

Что касается первого множителя в (3), он легко раскладывается следующим образом[2]:

$$\frac{\partial E}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} \cdot \frac{dy_k}{ds_k} \cdot \frac{\partial s_k}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} \cdot \frac{dy_k}{ds_k} \cdot w_{jk}^{(n+1)} \quad (5)$$

Здесь суммирование по k выполняется среди нейронов слоя $n+1$.

Введя новую переменную

$$\delta_j^{(n)} = \frac{\partial E}{\partial y_j} \cdot \frac{dy_j}{ds_j} \quad (6)$$

мы получим рекурсивную формулу для расчетов величин $\delta_j^{(n)}$ слоя n из величин $\delta_k^{(n+1)}$ более старшего слоя $n+1$.

$$\delta_j^{(n)} = \left[\sum_k \delta_k^{(n+1)} \cdot w_{jk}^{(n+1)} \right] \cdot \frac{dy_j}{ds_j} \quad (7)$$

Для выходного же слоя

$$\delta_i^{(N)} = (y_i^{(N)} - d_i) \cdot \frac{dy_i}{ds_i} \quad (8)$$

Теперь мы можем записать (2) в раскрытом виде:

$$\Delta w_{ij}^{(n)} = -\eta \cdot \delta_j^{(n)} \cdot y_i^{(n-1)} \quad (9)$$

Иногда для придания процессу коррекции весов некоторой инерционности, сглаживающей резкие скачки при перемещении по поверхности целевой функции, (9) дополняется значением изменения веса на предыдущей итерации

$$\Delta w_{ij}^{(n)}(t) = -\eta \cdot (\mu \cdot \Delta w_{ij}^{(n)}(t-1) + (1-\mu) \cdot \delta_j^{(n)} \cdot y_i^{(n-1)}) \quad (10)$$

где μ – коэффициент инерционности, t – номер текущей итерации.

Таким образом, полный алгоритм обучения НС с помощью процедуры обратного распространения строится так:

1. Подать на входы сети один из возможных образов и в режиме обычного функционирования НС, когда сигналы распространяются от входов к выходам, рассчитать значения последних.

$$s_j^{(n)} = \sum_{i=0}^M y_i^{(n-1)} \cdot w_{ij}^{(n)} \quad (11)$$

где M – число нейронов в слое $n-1$ с учетом нейрона с постоянным выходным состоянием $+1$, задающего смещение; $y_i^{(n-1)} = x_{ij}^{(n)}$ – i -ый вход нейрона j слоя n .

$$y_j^{(n)} = f(s_j^{(n)}), \text{ где } f() \text{ – сигмоид} \quad (12)$$

$$y_q^{(0)} = I_q, \quad (13)$$

где I_q – q -ая компонента вектора входного образа.

2. Рассчитать $\delta^{(N)}$ для выходного слоя по формуле (8).

Рассчитать по формуле (9) или (10) изменения весов $\Delta w^{(N)}$ слоя N .

3. Рассчитать по формулам (7) и (9) (или (7) и (10)) соответственно $\delta^{(n)}$ и $\Delta w^{(n)}$ для всех остальных слоев, $n=N-1, \dots, 1$.

4. Скорректировать все веса в НС

$$w_{ij}^{(n)}(t) = w_{ij}^{(n)}(t-1) + \Delta w_{ij}^{(n)}(t) \quad (14)$$

5. Если ошибка сети существенна, перейти на шаг 1. В противном случае – конец.

Сети на шаге 1 попеременно в случайном порядке предъявляются все тренировочные образы, чтобы сеть, образно говоря, не забывала одни по мере запоминания других. Алгоритм иллюстрируется рисунком 1.

Из выражения (9) следует, что когда выходное значение $y_i^{(n-1)}$ стремится к нулю, эффективность обучения заметно снижается. При двоичных входных векторах в среднем половина весовых коэффициентов не будет корректироваться, поэтому область возможных значений выходов нейронов $[0,1]$ желательно сдвинуть в пределы $[-0.5,+0.5]$, что достигается простыми модификациями логистических функций. Например, сигмоид с экспонентой преобразуется к виду

$$f(x) = -0.5 + \frac{1}{1 + e^{-\alpha x}} \quad (15)$$

для разделения множества входных образов, например, по двум классам достаточно всего одного выхода. При этом каждый логический уровень – "1" и "0" – будет обозначать отдельный класс.

Рассматриваемая НС имеет несколько недостатков. Во-первых, в процессе обучения может возникнуть ситуация, когда большие положительные или отрицательные значения весовых коэффициентов сместят рабочую точку на сигмоидах многих нейронов в область насыщения. Малые величины производной от логистической функции приведут в соответствие с (7) и (8) к остановке обучения, что парализует НС. Во-вторых, применение метода градиентного спуска не гарантирует, что будет найден глобальный, а не локальный минимум целевой функции. Эта проблема связана еще с одной, а именно – с выбором величины скорости обучения. Доказательство сходимости обучения в процессе обратного распространения основано на производных, то есть приращения весов и, следовательно, скорость обучения должны быть бесконечно малыми, однако в этом случае обучение будет происходить неприемлемо медленно. С другой стороны, слишком большие коррекции весов могут привести к постоянной неустойчивости процесса обучения. Поэтому в качестве η обычно выбирается число меньше 1, но не очень маленькое, например, 0.1, и оно, вообще говоря, может постепенно уменьшаться в процессе обучения.

Алгоритм подавления фона от адронных ливней тестировался на выходных данных эксперимента MAGIC.

В работе были использованы 3000 событий сгенерированных гамма-квантами и 3000 событий инициированных адронами.

Были использованы десять параметров изображения ливней:

параметры формы- width, length, параметры направления- alpha, dist и некоторые другие параметры описывающие особенности изображения ливня.

Данные были разделены на две части: 2000 событий соответствующих гамма-квантам были использованы для обучения сети(контрольные данные)и, остальные 1000-для тестирования функционирования сети(тестовые данные).Соответственно 2000 событий соотв. адронным ливням были использованы для обучения и 1000 для тестирования.

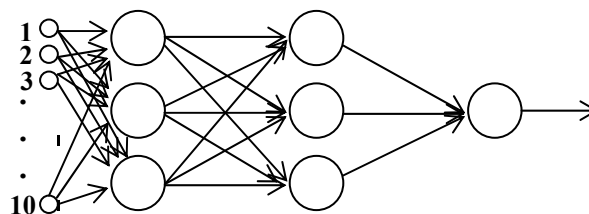


рис.8 нейронная сеть

Использовалась нейронная сеть состоящая из трех слоев и семи нейронов показанная на рис.8

Выход сети для каждого события представляет собой число в интервале [0;1]. Целью обучения является максимальное разделение значений выходов сети для событий двух классов, то есть желаемые выходы сети- 0 и 1 для событий соответствующих классов.

Реально выходы сети для события каждого класса представляют определенное распределение в интервале [0;1] (рис.9). Причисление события к первому или второму классу опопределяется так называемой "точкой решения", которая разделяет интервал на две части. Все события для которых выход сети находится в первой части интервала относятся к первому классу(гамма-кванты), остальные ко второму(адроны).

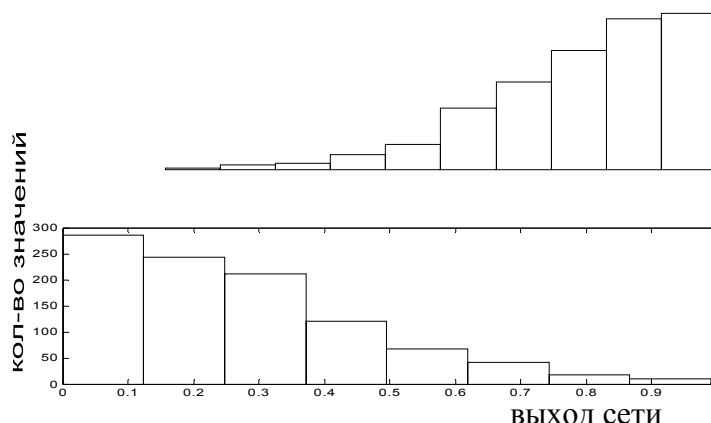


рис.9 Распределение выходов сети для гамма(верх.) и адронных(нижн.) ливней

Так, если необходимо уменьшить примесь адронных ливней "точку решения" сдвигают вправо, но при этом теряется и часть гамма-ливней. Наоборот, если

"точку решения" сместить влево увеличив количество зарегистрированных гамма-ливней, возрастает примесь адронных ливней.

Результаты классификации показаны на рис.10 , линия красного цвета соответствует контрольным данным, красная-тестовым.

Список использованной литературы:

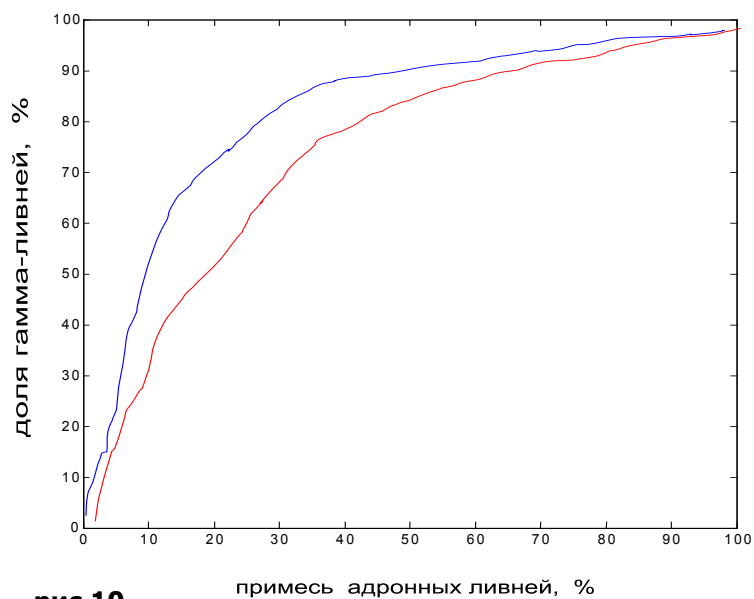


рис.10

1. B.Müller, J.Reinhardt Neural Networks, Springer-Verlag Berlin Heidelberg,1995

2. **С.Короткий , Нейронные сети**
3. **A.M. Hillas, Differences between Gamma-Ray and Hadronic showers.** Physics Department, University of Leeds, UK
4. **J.V. Jelley, T.C. Weekes, Ground-based Gamma-Astronomy ,** Sky & Telescope, Sept. 1995
5. **The MAGIC Telescope Project,** <http://hegra1.mppmu.mpg.de/MAGICWeb/>