



ГОСУДАРСТВЕННЫЙ НАУЧНЫЙ ЦЕНТР РОССИЙСКОЙ ФЕДЕРАЦИИ

ИНСТИТУТ ФИЗИКИ ВЫСОКИХ ЭНЕРГИЙ

ИФВЭ 96-75

ОМВТ

С.В. Клименко, Е.А. Слободюк, В.Д. Самойленко, Т.Л. Ухова

**ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ
В ФИЗИКЕ ВЫСОКИХ ЭНЕРГИЙ**

Протвино 1996

Аннотация

Клименко С.В. и др. Искусственные нейронные сети в физике высоких энергий: Препринт ИФВЭ 96–75. – Протвино, 1996. – 48 с., 5 рис., библиогр.: 68.

Данный обзор предназначен в первую очередь для специалистов, работающих в области экспериментальной физики высоких энергий (ФВЭ) и интересующихся новыми методами решения задач, для которых при традиционном подходе требуются суперресурсы. Обзор может быть полезен всем, желающим познакомиться с искусственными нейронными сетями (artificial neural networks – ANN) в конкретных приложениях.

Основной целью обзора является распространение новых методов и знаний, связанных с применением новых информационных технологий в ФВЭ. Авторы не ставили своей целью изложение основ ANN, а хотели привлечь внимание специалистов только к конкретным фактам успешного решения широкого спектра задач ФВЭ с помощью ANN.

Обзор состоит из введения, двух разделов и заключения. В первом разделе кратко изложены самые общие сведения о нейросетях, которые необходимы для понимания представленных в обзоре различных применений. Второй, основной раздел посвящён применению ANN в ФВЭ.

Abstract

Klimenko S.V. et al. Artificial Neural Networks for High Energy Physics: IHEP Preprint 96–75. – Protvino, 1996. – p. 48, figs. 5, refs.: 68.

This survey is destined for specialists in experimental high energy physics (HEP) who interested in novel techniques for challenge problem solving which demand superresources.

The major goal of the survey is dissemination of the new techniques and knowledge, concerned with the application of novel information technology in HEP. The authors did not intend to present the fundamental of ANN, just they wanted to attract readers to concrete facts of successful HEP problem solving with using ANN.

The survey includes two sections. The first section presents a general information about ANN being needed for understanding the following applications. The second section denotes to the applications of ANN in HEP.

Содержание

Введение	2
1. Искусственные нейронные сети	4
1.1. Персептрон. Метод обратного распространения ошибки	4
1.2. Самоорганизующиеся сети	10
1.3. Сети с обратной связью	12
1.4. Реализация ANN в виде СБИС	14
2. Применение нейронных сетей в ФВЭ	16
2.1. Информационная структура эксперимента в ФВЭ	17
2.2. Применение в триггерах	20
2.2.1. Триггер на b -кварк	22
2.3. Классификация и идентификация событий	23
2.3.1. Разделение кварк-глюонных струй	23
2.3.2. Идентификация тяжелых кварков	27
2.3.3. Идентификация распадов τ -лептона	27
2.3.4. Поиск хиггсовского бозона H	28
2.3.5. Идентификация системы частиц в конечном состоянии	30
2.3.6. Идентификация частиц по энергетическим потерям	33
2.3.7. Идентификация частиц в детекторе типа RICH	34
2.3.8. $n - \gamma$ разделение в нейтронной спектрометрии	34
2.4. Реконструкция треков	36
2.5. Вычисление массы	37
2.6. Реконструкция продольного фазового пространства пучка заряженных частиц по спектру синхротронного излучения	39
2.7. Обработка данных времяпроекционной камеры	41
2.8. Фитирование профиля пучка	43
Заключение	43
Список литературы	44

Введение

Нейронные сети представляют собой новую технологию вычислений, которая базируется на создании математических структур, способных к обучению. Эта технология является результатом фундаментальных исследований в области искусственного интеллекта при моделировании обучения в системах, подобных мозгу или нервной системе. Несмотря на искусственность и академичность подхода результат исследования имеет очень хороший практический выход.

Нейронные сети с их замечательной способностью получать результаты из сложных или неопределённых данных могут быть использованы для распознавания образов и определения тренда даже в тех случаях, когда становятся бессильными любые другие компьютерные технологии. Тренированные нейронные сети могут выступать в роли “серьёзных экспертов” для той категории проблем, на которую они были “натасканы”. Эти “эксперты” могут успешно решать задачи в совершенно новых ситуациях из заданной предметной области, которые (задачи) требуют ответа на вопросы “что, если”.

Нейронные сети зарекомендовали себя наилучшим образом при решении многих прикладных проблем, в том числе в таких нетрадиционных для применения компьютеров областях, как

- исследование и прогнозирование рынка,
- управление процессом производства,
- управление риском и многие другие.

Исследования на переднем крае фундаментальных наук всегда были пробным камнем высоких технологий. Поэтому проблемы применения искусственных нейронных сетей для задач физики высоких энергий являются чрезвычайно актуальными. Об этом свидетельствует хотя бы такой факт, что если Вы воспользуетесь замечательной системой AltaVista глобального поиска в среде информационных ресурсов Интернет [68] с ключевыми словами “neural network” + “high energy physics”, то найдёте более 10 000 документов, в которых первый ключ встречается 100 000 раз, а второй — 30 000 раз.

Данный обзор представляет собой компиляцию самых общих сведений о нейронных сетях и новейших достижений по применению этого величайшего изобретения в традиционно самой сложной и ресурсоёмкой области фундаментальных исследований — в физике высоких энергий.

Обзор состоит из двух разделов. В первом кратко представлены основные сведения о нейронных сетях, необходимые для понимания последующего материала о конкретных применениях ANN в физических задачах. Во втором, основном разделе обзора сначала приведены сведения об информационной структуре эксперимента в физике высоких энергий и названы основные элементы экспериментальных установок, для которых целесообразно применение нейросетевого подхода при обработке данных.

Чрезвычайно интересным и эффективным является применение ANN в процедуре предварительного отбора событий (выработки триггера), которая требует “мгновенной” обработки огромного количества разнородной информации, для определения принадлежности физического события к избранному классу. Далее приведены результаты использования нейросетей для широкого класса задач классификации и идентификации. Ряд представленных применений в задачах реконструкции показывает, как традиционные вычислительные процедуры эффективно решаются с помощью невычислительных (нейросетевых) методов. В заключении сформулированы выводы и обозначены перспективы применения ANN в будущих всёусложняющихся задачах экспериментальной физики высоких энергий.

Для начала, не вдаваясь в подробности биофизики нейрона [25], рассмотрим наиболее часто используемую модель, отражающую его основные свойства и представляющую нейрон как пороговый элемент:

$$v_i = g \left(\sum_j w_{ij} v_j + \theta_i \right), \quad (1)$$

где v_i, v_j — это значения функции состояния для i -го и j -го нейронов соответственно; w_{ij} — веса, выражающие силу связи нейрона i с нейроном j ; g — функция перехода; θ_i — порог, соответствующий мембранному потенциалу биологического нейрона. Веса могут быть положительными (возбуждающими) и отрицательными (тормозными). Суммирование в (1) ведется по всем нейронам, связанным с нейроном i . Веса и функция перехода представляют собой локальные параметры, описывающие данный нейрон. На практике используют модели нейронов со значениями v_i , лежащими в интервалах $[-1, 1]$ (или $[0, 1]$), а также простые двоичные нейроны со значениями $v_i = \{-1, 1\}$ (или $\{0, 1\}$). Двоичные нейроны обычно обозначаются s_i .

Для различных задач используются разные функции перехода. Наиболее простая из них — функция Хевисайда.¹

$$g(x) = \begin{cases} 1, & \text{если } x \geq 0 \\ 0, & \text{в противном случае.} \end{cases} \quad (2)$$

Наиболее общей функцией *активации нейрона* является сигмоида (sigmoid-shaped), в качестве которой часто берётся гиперболическая тригонометрическая функция в форме

$$g(x) = \tanh(ax). \quad (3)$$

Наиболее типичным примером сигмоиды является *логистическая функция*

$$g(x) = 1/(1 + e^{-ax}), \quad (4)$$

где a есть *параметр наклона* сигмоидной функции.

Существуют и другие модели нейронов.

¹Модель двоичного нейрона с такой функцией перехода была предложена Маккаллохом и Питтсом еще в 1943 году (см. работу [42]) и получила название формальной модели.

1. Искусственные нейронные сети

Здесь кратко будут представлены основные простейшие модели нейронных сетей. Для более подробного знакомства с нейронными сетями см. работы [5], [47], [52].

Одним из важнейших достоинств нейронных сетей является их обучаемость. В простейшем случае проблема обучения сети сводится к вопросу выбора весов. Не обладая достаточной информацией для выбора весов *a priori*, используют их итеративную настройку. Способы обучения таких сетей делятся на два класса: обучение при заданной классификации (с учителем — supervised learning) и обучение с неизвестной классификацией (без учителя — unsupervised learning). Обучение при заданной классификации основано на непосредственном сравнении выхода сети с известным ответом. (Отметим, что для обучения сети необходимо иметь адекватные примеры).

Конкретное использование сети делится на два этапа: первый — обучение сети на примерах (training), второй — работа сети на данных, которые она до этого никогда не видела (generalization — процедура обобщения). При обучении с неизвестной классификацией нет примеров, определяющих цель обучения. Информация, требуемая для обучения, может быть получена лишь из корреляции входных данных. Сеть создает категории этих взаимосвязей, и её выходной сигнал будет соответствовать найденной классификации.

Архитектура простейших нейронных сетей может быть представлена двумя основными видами: сети с прямой связью (feed-forward, сокращённое название FFNN) и сети, имеющие ещё и обратную связь (feed-back, сокращённо FBNN). В FFNN сети входные сигналы проходят снизу вверх, слой за слоем, используя правило перехода. В FBNN сети все узлы (нейроны) взаимосвязаны между собой, а синапсы (связи) имеют два направления.

1.1. Персептрон. Метод обратного распространения ошибки

Упрощённой моделью FFNN является простой персептрон (simple perceptron), он состоит из двух слоёв, являющихся входом и выходом. На этой модели хорошо работает алгоритм с заданной классификацией, обучающий сеть на примерах $\vec{x}^{(p)} \rightarrow \vec{t}^{(p)}$. Этот алгоритм основан на минимизации функции суммарной квадратичной ошибки

$$E = \frac{1}{2} \sum_p \sum_i (o_i^{(p)} - t_i^{(p)})^2 \quad (5)$$

по отношению к весам, где $o_i^{(p)}$ — i -ый выход сети, закончившей работу на входном образце $\vec{x}^{(p)}$,

$$o_i^{(p)} = \sum_j w_{ij} x_j^{(p)}, \quad (6)$$

$t_i^{(p)}$ — желаемый выход (известный правильный отклик сети на этом образце).

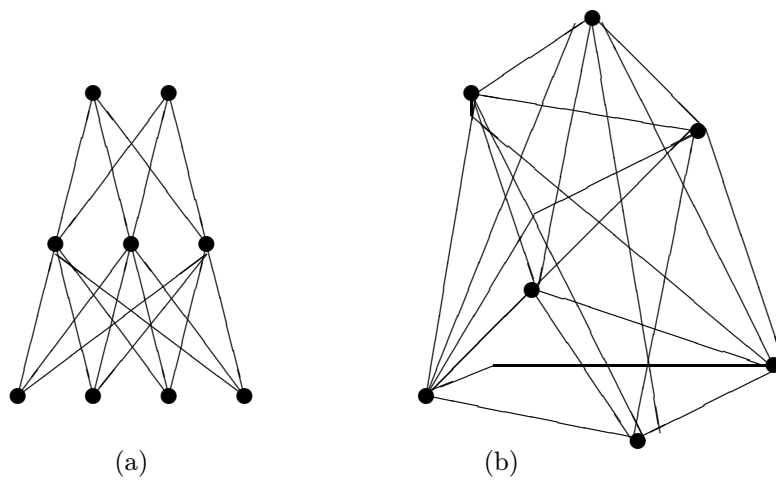


Рис. 1. (a) - сеть с прямой связью; (b) - сеть, имеющая обратную связь.

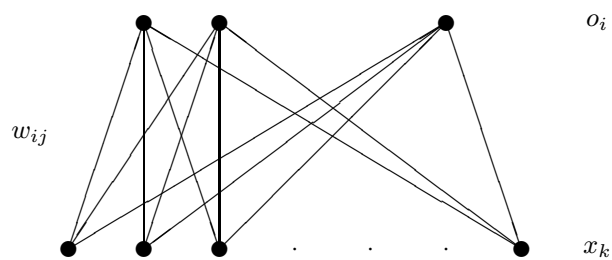


Рис. 2. Простой персептрон.

Будем считать, что функция перехода непрерывна и задается формулой (3). Для тренировки перцептрона используется метод градиентного спуска. Веса изменяются следующим образом:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}, \quad (7)$$

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}, \quad (8)$$

где η — параметр, от которого зависит скорость обучения, ($\eta < 1$).

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_i} \frac{\partial o_i}{\partial w_{ij}}, \quad (9)$$

$$o_i = g\left(\sum_{j=0} w_{ij} x_j\right) = g(\vec{w}_i \vec{x}), \quad (10)$$

$$\Delta w_{ij} = -\eta \delta_i x_j, \quad (11)$$

где

$$\delta_i = (o_i - t_i) g'(\vec{w}_i \vec{x}). \quad (12)$$

После каждого приращения весов берется новый образец $\vec{x}^{(p)}$ из тренируемого ряда. Когда приращение весов Δw_{ij} становится равным нулю (с некоторой точностью), веса замораживаются и сеть готова работать на данных, которые она до этого никогда не встречала.

Недостатком простого перцептрона является его линейность в том смысле, что он не способен отделять в N -мерной входной сфере классы, разделенные между собой $(N-1)$ -мерной гиперплоскостью. Так, например, он не решает XOR проблему (отделение разных по значению точек логической функции “исключающее или” (exclusive-or)). Преодолеть это препятствие можно двумя способами: первый — усложнить функцию перехода, второй более сильный и общий — ввести промежуточные (hidden — скрытые) слои между входным и выходным слоями. Такую сеть обычно называют многоуровневый перцептрон (multilayer perceptron (MLP)). Узлы промежуточных слоёв обозначаются обычно h_j .

Градиентный метод, обобщённый для сети с промежуточными слоями, получил название back-propagation (BP). BP часто называют методом обратного распространения ошибки. В BP появляется опасность попадания в локальный минимум и в противоположность простому перцептрону нет гарантии, что глобальный минимум будет достигнут. Время сходимости задачи быстро растет с ростом её размерности. Существуют различные варианты и расширения этого метода.

В качестве меры возможной ошибки может быть взята любая дифференцируемая функция, имеющая минимум в точке $o_i = t_i$. Можно использовать, например, ошибку перекрёстной энтропии (Cross-entropy):

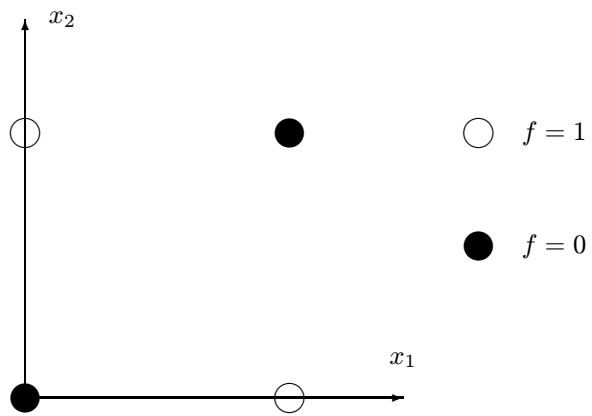


Рис. 3. Функция “исключающее или”.

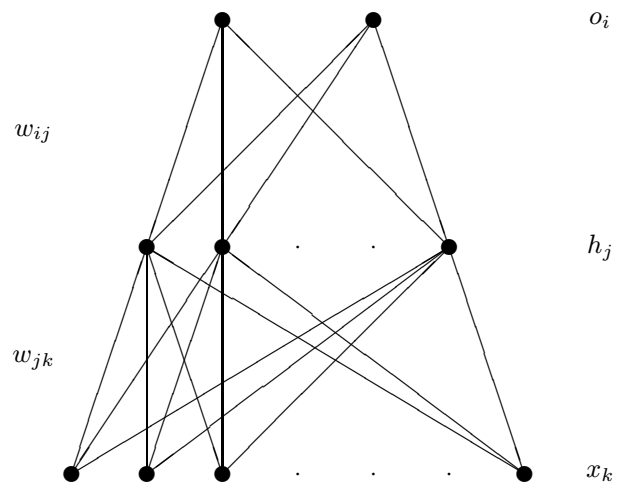


Рис. 4. FFNN с одним промежуточным слоем.

$$E = - \sum_p \sum_i [t_i^{(p)} \log o_i + (1 - t_i^{(p)}) \log(1 - o_i)] \quad (13)$$

или меру Кульбака(Kullback):

$$E = \sum_p \sum_i t_i^{(p)} \log \frac{t_i^{(p)}}{o_i} \quad (14)$$

В этих случаях множитель g' , содержащийся в формуле перевычисления весов, исчезает.

BP в основной своей форме имеет два основных параметра: температуру T и обучающую скорость η . В течение обучения сеть проходит через различные точки энергетической поверхности, для каждой из которых оптимальными будут свои значения обучающих параметров.

На практике применяют так называемый метод “отжига” (annealing), использующий медленное понижение температуры T в процессе обучения, т.е. понижают среднюю энергию нестабильного состояния относительно локального минимума, что в большинстве случаев позволяет попасть в глобальный минимум.

Слишком большие значения η вызывают колебания при тренировке сети. В конце обучения η должна приближаться к нулю для того, чтобы сеть попала в стабильное состояние. Оптимальная обучающая скорость для веса w_{ij} :

$$\eta_{ij} \propto 1/\mu_j \quad ,$$

где μ_j — число нейронов, связанных с нейроном j .

Чтобы избежать колебаний, весам придают некоторую инерцию, добавляя дополнительный член:

$$\Delta w(t+1) = -\eta \frac{\partial E}{\partial w} + \alpha \Delta w(t). \quad (15)$$

Параметр α , называемый моментным, должен лежать между 0 и 1; он часто устанавливается очень близко к 1.

Существуют методы со стохастическими элементами, имеющие целью избежать попадания в локальный (побочный) минимум энергии. Добавление шума к входным образцам улучшает процедуру обобщения. Но слишком большой шум действует на работу сети разрушающим образом.

Значение первоначальных весов также влияет на работу сети. Веса выбираются случайным образом из интервала $[-w_0, +w_0]$. Начальное их изменение должно быть большим. Согласно “правилу большого пальца” первоначальные веса должны быть $w \approx w_0 / \max_j \mu_j$, где $w_0 \approx T/10$. $\Delta w_{ij} \propto g'()$, а определенная часть g' растет с T , поэтому очень низкие T будут разрушать веса. $T = 1$ обычно хороший выбор.

Важным вопросом является выбор структуры сети (числа слоев, промежуточных нейронов, степень их взаимосвязанности). Желательно использовать как можно меньшее количество весов, насколько это возможно для успешной работы сети. Существуют много способов построения сети со структурой, близкой к оптимальной.

Те, которые получили название “start big and remove”, начинают с большого числа весов, а затем устраняют такие веса, в которых нет сильной потребности для достижения успеха. Важность веса определяется путем оценки чувствительности ошибочной функции к устранению данного веса (здесь используются обычно эвристические методы). Можно также модифицировать обучающий алгоритм так, чтобы после обучения лишние связи обнулялись. Например, в методе “распада весов” (weight decay) редко изменяющимся весам позволено разрушаться в соответствии с формулой

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} - \varepsilon w_{ij}, \quad (16)$$

где ε — параметр распада, обычно малая величина, $O(10^{-4})$. Это соответствует добавлению дополнительного члена к функции энергии:

$$E \rightarrow E + \frac{\varepsilon}{2\eta} \sum_{ij} w_{ij}^2. \quad (17)$$

Другие способы (“start small and add”) построения сети начинают свою работу от простой сети и потом добавляют связи для уменьшения ошибки. Существуют алгоритмы, основанные на разделении сети на подсети, каждая из которых может изменяться путём приращения весов.

Примером сети, включающей в себя оба эти направления может служить сеть GAL (Grow and Lean) в двух своих модах “sleep” и “awake” (см. работу [3]). Эта сеть учится на ассоциациях благодаря способности к приращению и использованию локальных представлений.

Для любой задачи, использующей FFNN, достаточно не более двух промежуточных слоёв.

Необходимо, чтобы обученная сеть хорошо работала в процедуре обобщения, т.е. на данных, которых она до этого не встречала. С увеличением времени тренировки эффективность сети возрастает до некоторого уровня, дальнейшее же увеличение времени тренировки ведет к ухудшению способности сети к обобщению (наступает переобучение).

Наилучшее обобщение дается пределом Байеса, представляющим собой минимальное перекрытие распределений, вероятностей попадания событий в разные классы (см. работу [20]). Использование слишком большого количества весов ведет к ухудшению процедуры обобщения. Для FFNN ошибка обобщения $\epsilon \sim O(N_w/N_p)$, где N_w — число весов, N_p — число образцов, необходимых для тренировки сети. Согласно “правилу большого пальца” число образцов должно быть по крайней мере в 10 раз больше числа весов в сети, $N_p \geq 10N_w$.

BP на первый взгляд может показаться очень сильным. Но он хорошо справляется с тренировкой сети, имеющей лишь небольшое количество входов. Это обусловлено следующими причинами:

1) настройка весов в *BP* предполагает бесконечно малое их изменение, что влечёт за собой увеличение числа образцов, представленных для обучения сети и времени обучения,

2) для достижения хорошего результата работы сети на тестируемом ряде число образцов тренирующего ряда должно быть значительно больше числа весов в сети.

К тому же существует еще одно неудобство метода: тренирующий ряд не может быть изменен. Изменение даже одного сигнала во входном образце требует повторения процесса обучения сети на всем тренирующем ряде.

Из вышеперечисленных ограничений метода следует, что *BP* вряд ли подходит для распознавания сложных объектов. Плохо поставленные задачи, а также задачи, не имеющие очевидных правил решения, выходят за пределы возможностей этого метода. Различные модификации метода улучшают его, но не снимают упомянутых ограничений.

При применении *BP* для уменьшения тренирующего ряда часто используют предварительную обработку входных данных. Такая обработка данных может быть осуществлена с помощью самоорганизующихся сетей.

1.2. Самоорганизующиеся сети

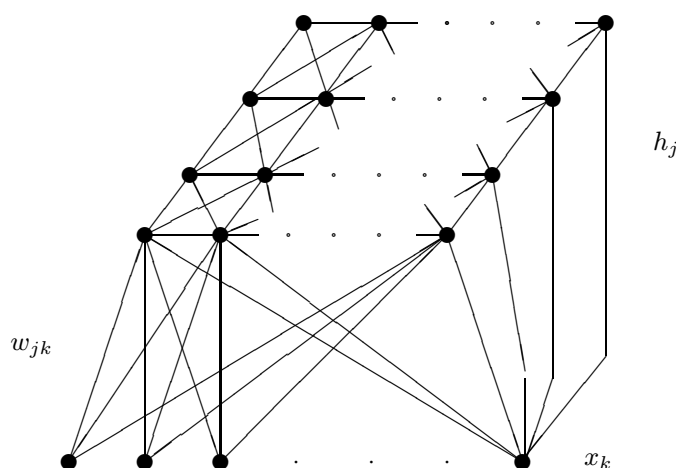


Рис. 5. Самоорганизующаяся сеть.

Сети с прямой связью FFNN, основанные на способе обучения с неизвестной классификацией, т.е. не использующие какого-либо обучения на заранее известных примерах, называются самоорганизующимися (self-organizing — SO). Такая сеть должна открывать для своих образцов наиболее часто встречающиеся признаки, регулярности, корреляции и кодировать их на выходе. SO-сеть может быть таким образом использована для выделения основной структуры данных. Она содержит обычно только один слой кроме входного, нейроны этого слоя носят название нейронов характерных признаков (feature units) и обозначаются h_j . Нейроны характерных признаков могут быть взаимосвязаны. В этом случае между ними существует как прямая, так и обратная связь.

По сложности SO-сети делятся на два класса: “победитель берет все” (competitive) и коллективные версии (collective SO).

Примером коллективной версии SO может служить VQ (vector quantization). Каждый узел h_j ставят в соответствие весовым векторам

$\vec{w}_j = (w_{j1}, w_{j2}, \dots, w_{jN})$, имеющим ту же самую размерность, что и входной вектор

$\vec{x}^{(p)} = (x_1^{(p)}, x_2^{(p)}, \dots, x_N^{(p)})$, где p — порядковый номер образца во входном ряде, а w_{jk} — весовой вектор, связывающий вход x_k с нейроном h_j . Для каждого входного образца $\vec{x}^{(p)}$ вычисляется расстояние между весовым вектором и образцом. Наиболее близкий образец становится “победителем”:

$$h_m = \min_j (|\vec{w}_j - \vec{x}^{(p)}|). \quad (18)$$

Победивший вектор \vec{w}_m двигается тогда ближе к входному образцу $x^{(p)}$:

$$\Delta w_m = \eta (\vec{x}^{(p)} - \vec{w}_m), \quad (19)$$

где η — обучающая скорость. Это соответствует градиентному спуску на ошибочной функции

$$E^{(p)} = \frac{1}{2} \sum_{p \in M} (\vec{x} - \vec{w}_m)^2, \quad (20)$$

где M — ряд входов, имеющих нейрон h_m в качестве победителя. Весовые векторы, таким образом, будут сходиться к $\langle \vec{x} \rangle_M$, центру масс кластера M . Нейроны характерных признаков будут разделять входную область данных на некоторое число подобластей, заданных многогранниками. В начале обучения обучающая скорость η должна быть большая, чтобы могли быть достигнуты большие изменения Δw_j , а затем уменьшена для того, чтобы сеть попала в стабильное состояние. Как и в *BP* первоначальные веса в VQ инициализируются случайным образом.

Для задач классификации также может быть применен алгоритм LVQ (learning vector quantization), представляющий собой VQ, к которому добавлено обучение, использующее примеры. Здесь происходит движение весового вектора от нежелательных входных образцов путем изменения знака η в уравнении (19) всякий раз, когда входные данные не классифицируются верно.

LVQ более эффективен и быстр для задач малых размерностей в отличие от *BP*, который дает лучшие результаты для больших размерностей, так как LVQ заполняет входную область многогранниками, а *BP* делит её, используя гиперплоскости, что для задач больших размерностей более экономично.

Случай, когда нескольким нейронам позволено адаптироваться к данному входу $\vec{x}^{(p)}$, получил название коллективной версии SO.

Метод “отображения характерных признаков” (feature mapping) является одной из версий VQ. Этот метод создает топологически корректное отображение входной области на “карте характерных признаков” — “feature map”. Для нейронов h_j геометрическая топология должна быть определена заранее. Обычно она представляет собой двумерную решетку.

На карте характерных признаков топологический порядок достигается благодаря модернизации весов соседей победителя:

$$\Delta \vec{w}_j = \eta \Lambda_{(j,m)} (\vec{x}^{(p)} - \vec{w}_j), \quad (21)$$

где $\Lambda_{(j,m)}$ — функция близости (соседства). Она является функцией (евклидового) расстояния d_{jm} между нейронами характерных признаков j и m :

$$\Lambda_{(j,m)} = \begin{cases} 1, & \text{если } d_{j,m} < \lambda \\ 0, & \text{в противном случае,} \end{cases} \quad (22)$$

λ — параметр, называемый шириной и определяющий пластичность сети.

Если значение параметра λ большое, то нейроны характерных признаков сильно связаны между собой и сеть жесткая. Если λ — нуль, эти нейроны полностью свободны и им позволено сходиться к центрам соответствующих им кластеров, сеть в этом случае мягкая.

В начале обучения величина параметра λ должна быть большая, порядка размера карты характерных признаков, для того, чтобы сеть нашла основную локализацию данных. В процессе работы сети λ уменьшают, чтобы получить более детальную картину распределения данных.

Весы наиболее близких к победителю соседей изменяются в том же направлении, что и вес победителя. В конце работы сети ближайшие нейроны будут иметь одинаковые весовые векторы, а результирующая сеть будет топологически корректна (если это возможно), т.е. соседним точкам во входной сфере будут соответствовать соседние точки на карте характерных признаков.

Не существует доказательств сходимости алгоритма для N -мерной карты, но может быть показано, что для одно- и двухмерной карты существует сходимость к равновесному состоянию, если $\eta(t) \rightarrow 0$ при $t \rightarrow \infty$.

Недостатком SO-сетей является то, что они требуют априорной информации (в форме представления весов и управляющих параметров).

1.3. Сети с обратной связью

В этом разделе будут рассмотрены основные свойства сетей с обратной связью (feed-back) NN — FBNN. Наиболее распространённой моделью такой сети, представляющей собой ассоциативную память, является модель Хопфилда (Hopfield model). Она использует идеи статистической теории систем спинов и идеи Хебба о процессах обучения в мозге.

Модель Хопфилда основана на энергетической функции

$$E = -\frac{1}{2} \sum_{i \neq j} \sum_j w_{ij} s_i s_j \quad (23)$$

с двоичными нейронами $s_i = \pm 1$ и симметричными связями ($w_{ij} = w_{ji}$). Динамическое изменение энергии (23) обуславливается изменением s_i в соответствии с

формулой

$$s_i = \operatorname{sgn} \left[\sum_{j \neq i} w_{ij} s_j \right]. \quad (24)$$

Достижение ассоциативности означает, что для сети разные входные векторы, если они достаточно одинаковы, вызывают один и тот же отклик. Для этого необходимо правильно выбрать веса. Затем нейроны устанавливаются в стабильное состояние и решение задачи будет задаваться конфигурацией $\vec{s} = (s_1, s_2, \dots)$, дающей минимум энергии.

Выбор весов происходит один раз по правилу Хебба на ряде образцов $\vec{x}^{(p)}$:

$$w_{ij} = \sum_{p=1}^{N_p} x_i^{(p)} x_j^{(p)}, \quad (25)$$

т.е. веса возрастают между нейронами, чьи активности коррелируются и уменьшаются между антикоррелирующими нейронами. Веса могут принимать любой знак, что приводит к интересному энергетическому ландшафту. Образцы (входные конфигурации \vec{s}) оказываются расположенными в локальных минимумах энергии.

После вычисления значений весов по правилу Хебба выбирают начальное состояние $\vec{s} \approx \vec{x}^{(p)}$. Затем берут по одному нейрону, модернизируют их в соответствии с формулой (24) и вычисляют $\Delta \vec{s} = \vec{s}_{[i+1]} - \vec{s}_{[i]}$, где номер в квадратных скобках означает номер итерации. При этом выбор нейронов осуществляется случайным образом. Процедуру повторяют до тех пор, пока $\Delta \vec{s}$ не будет равна 0, т.е. не будет достигнуто стабильное состояние, являющееся решением задачи.

Наличие стабильного состояния гарантируется симметричностью связей.

Основное преимущество “памяти Хопфилда” — ассоциативность и надёжность. Если в представленном сети образце $x^{(p)}$ разрушен некоторый бит, с помощью уравнения (24) он будет восстановлен (сеть начнет “скользить” вниз по направлению к правильному образцу, находящемуся в локальном минимуме энергии E). Так же, если потерян некоторый вес w_{ij} , сеть будет работать хорошо, так как память распределена по всем весам в противоположность BP , где энергия является функцией весов.

Однако модель Хопфилда не устраняет духов — поддельных (spurious) состояний, и не может быть применена к трудным задачам распознавания, так как обучающее правило Хебба не рассчитано на наличие скрытых (hidden) узлов. Эти проблемы решает машина Больцмана (Boltzmann machine — BZ) [47], [24], [52], представляющая собой расширение сети Хопфилда путем включения скрытых узлов.

Сеть называется BZ потому, что вероятность состояний системы дается с помощью больцмановского распределения статистической механики. Основная проблема для этой сети — найти правильные связи к скрытым узлам, не имея обучающих образцов, которые должны представлять эти узлы.

VZ — алгоритм очень медленный, но его детерминистическая версия (MFT — Mean Field Theory — теория среднего поля) — довольно быстрая (см. работы [47], [49]).

Модели ассоциативной памяти используются при распознавании и реконструкции образов, при поиске в базах данных, восстановлении библиографической информации по частичным ссылкам. FBNN нашли свое применение в нахождении приближенных решений трудных задач оптимизации, являющихся NP-полными. Точное решение таких задач требует $n!$ вычислений при размерности проблемы равной n . MFT был с успехом применен к проблеме разделения графа и к задаче коммивояжера (см. работы [21], [50]).

1.4. Реализация ANN в виде СБИС

Благодаря большим успехам микроэлектроники стало возможным реализовать ANN в виде СБИС (VLSI) с параметрами, достаточными для применения в физическом эксперименте. Преимуществом таких СБИС является быстрая параллельная обработка сигналов и выполнение сложных задач распознавания образов.

Наибольшее распространение в экспериментах физики высоких энергий получила аналоговая СБИС ETANN 80170NX фирмы INTEL [58], содержащая 64 входа, 16 внутренних уровней и 64 нейрона (пороговый усилитель с сигмоидной передаточной функцией). Каждый вход соединён с 64 синапсами. ETANN может образовывать ANN со следующей архитектурой:

- 64 входных нейрона и 64 выходных нейрона,
- 64 входных нейрона, 64 скрытых нейрона и 64 выходных нейрона,
- 128 входных нейрона и 64 выходных нейрона.

Передаточная функция нейрона в СБИС близка к сигмоиде

$$f(x) = \frac{2}{1 + e^{-Gx}} - 1, \quad (26)$$

где усиление G (*параметр наклона*) может изменяться с помощью V_{gain} . Усиление передаточной функции определяет чувствительность нейрона. Низкое значение усиления позволяет интерпретировать выход нейрона как аналоговый, а высокое — как цифровой. Максимальное значение выхода нейрона определяется напряжением V_{ref0} . Веса ограничены интервалом $[-2.5, 2.5]$. Скорость прохождения сигнала по одному слою зависит от усиления и в среднем $\simeq 1,5$ мкс, что и определяет быстродействие. Точность выполнения операций примерно эквивалентна 6 битам, быстродействие — $1,3 \cdot 10^9$ переключений/секунду. Обучение выполняется методом Back Propagation с помощью Intel Neural Network Training System (INNTS) [59], работающей на компьютере i486. Применяемое системное окружение представляет собой специальную версию пакета DynaMind. Обучение выполняется до получения приемлемого уровня ошибки выхода сети, и после достижения удовлетворительной работы веса загружаются в СБИС. Для реальной работы такого обучения недостаточно, так как программа симуляции не может точно смоделировать аналоговую

работу СБИС, и, например, не отслеживает флуктуации в передаточной функции каждого нейрона. Поэтому следующий этап обучения представляет собой так называемый *chip-in-loop* (CIL) training, когда после каждого цикла веса записываются в СБИС и выход сети непосредственно используется в процессе обучения.

Поскольку ETANN представляет собой аналоговую СБИС, то для ее надежной работы важны стабильные внешние условия. Специально сконструированный для этого модуль обеспечивает низкую пульсацию источника питания $\delta V < 5$ мВ (напряжение питания 5 В) и температурную стабильность при 18°C $\delta T < 1^\circ\text{C}$ (потребляемая мощность ETANN 5 Вт). Конструктивно СБИС расположены на плате в стандарте VME.

Фирмой Siemens был разработан СБИС MA16 [60] как прообраз нейровычислительного устройства. MA16 представляет собой программируемый, каскадируемый процессор для векторных и матричных операций. Он поддерживает на аппаратном уровне следующие операции:

- *матричное умножение,*
- *матричное сложение/вычитание,*
- *нормировка результата,*
- *вычисление векторной нормы (метрики $L1$ и $L2$),*
- *вычисление векторного расстояния (мера Манхэттанна, геометрическое расстояние).*

СБИС содержит четыре идентичных процессорных элемента, работающих параллельно. Входные данные имеют точность 16 бит, тактовая частота 50 МГц. Для операций матричного умножения/сложения скорость вычислений достигает $8 \cdot 10^8$ операций/секунду. Программное обеспечение работает в среде UNIX/XWIND и реализовано на C++. Нейронная сеть тоже описывается на C++, или может вводиться интерактивно с помощью графического интерфейса типа OSF/Motif, что позволяет визуализировать конфигурацию СБИС после отображения на нее структуры сети. Средства тестирования и эмуляции развиты очень хорошо. С 1995 г. СБИС MA16 является коммерчески доступным продуктом.

Несомненный интерес вызывает опытная разработка фирмы Bellcore — CLNN32/ CLNN64 [7], [61]. CLNN32 состоит из 32 нейронов и 496 двунаправленных адаптивных синапсов. CLNN64 содержит только 1024 адаптивных синапсов. В наборе CLNN32/CLNN64 все нейроны взаимосвязаны, так что любая топология сети отображается подбором синапсов. Динамика сети полностью аналоговая, но значения синапсов хранятся/обновляются в цифровом виде с точностью 5 бит. На аппаратном уровне реализовано обучение сети — подбор весов происходит по алгоритму обучения машины Больцмана или Mean Field. Внутри также имеется некоррелированный генератор шума (32 канала), используемый при обучении по методу машины Больцмана. CLNN32 может быть использован независимо или совместно с CLNN64 для построения более сложной архитектуры сети. Производительность достигает 10^8 переключений/секунду (при работе с CLNN64 удваивается). Для CLNN32 это означает, что $\simeq 10^5$ 32-бит образцов/секунду или 32

аналоговых канала (с полосой пропускания 50 кГц) могут быть использованы для быстрого распознавания/обучения. Время распространения для одного слоя нейронов < 1 мкс. “Охлаждение” (по методу Больцмана) или MF обучение требует 10 – 20 мкс. По сравнению с ETANN СБИС CLNN32 имеет следующие очевидные преимущества:

- *быстрое обучение (микросекунды по сравнению с часами при СІІ процессе);*
- *эффективный алгоритм обучения Больцмана, обеспечивающий быстрое нахождение “почти оптимального” решения;*
- *простые и быстрые процедуры чтения/записи весов, выполняемые в цифровом виде, что значительно увеличивает скорость обмена между сетевым сервером и клиентами в сети;*
- *легкая каскадируемость.*

В университете Твенте совместно с DESY разработана, и в настоящее время работает в эксперименте H1, аналоговая СБИС NeuroClassifier [62]. Она содержит входной слой (70 входов, полоса пропускания до 4 Гбайт/сек), 6 внутренних слоёв и один выходной нейрон. Точность аналогового умножения 5 бит, время решения всего лишь 20 нс, что позволяет использовать NeuroClassifier в триггере первого уровня. Эквивалентное быстроедействие $\simeq 2 \cdot 10^{10}$ переключений/секунду.

Иногда нейронную сеть реализуют на БИС широкого применения [67]. Такой подход не позволяет уместить всю сеть на одну-две платы, но зато можно оптимизировать структуру под узкоспециальную задачу с хорошими результатами.

В последующих разделах будут рассмотрены конкретные примеры использования СБИС в физическом эксперименте.

2. Применение нейронных сетей в ФВЭ

Как уже упоминалось во введении, физика высоких энергий может служить эффективным пробным камнем для ANN. Нейронные сети хорошо подходят для решения задач, для которых нет строгих методов решения. К таким задачам прежде всего относятся задачи распознавания образов, идентификации и классификации.

Для тех задач, где имеются явные правила решения какой-либо проблемы, нейронная сеть может помочь избежать утомительной разработки программ или специализированной электроники, гарантируя быстроту и надёжность решения. Сюда относятся задачи выработки триггера и геометрической реконструкции.

Применение нейронных сетей основано на таких свойствах, как обучаемость, быстроедействие, нелинейность входа от выхода, терпимость к шуму, возможность использовать ассоциативную память, способность адаптироваться к новым условиям.

В зависимости от типа задачи для построения солвера (solver — решатель) выбирают ту или иную нейронную сеть. Ф. Фогелман-Соули (F. Fogelman-Soulie) в [22] приводит следующую таблицу применения нейронных сетей:

Тип обучения	С известной классификацией (supervised)	С неизвестной классификацией (unsupervised)
Тип проблемы		
Ассоциативная память	Linear Associative Memory Hopfield	MLP(auto-encoding mode)
Классификация Предсказание Интерполяция	Adaptive-Perceptron MLP-TDNN LVQ- RBF	
Квантизация (Quantization)	LVQ, Nestor	Topological Map ART

Описание алгоритмов TDNN (Time Delay Neural Network), RBF (Radial Basis Function), ART (Adaptive Resonans Theory) можно найти в [24].

Наибольший эффект при использовании нейронных сетей наблюдается в случаях, когда необходимо довольно быстро получить приближенное решение какой-либо конкретной задачи. Таких задач также немало в физике высоких энергий.

На сегодняшний день разработано много программных пакетов, моделирующих различные нейронные сети. Информацию о моделирующих ANN пакетах можно найти в работе [22]. В основном они написаны на языке C. В ФВЭ используются также программные пакеты, написанные на Фортране, такие как Fermilab Simulator и JETNET (см. работы [63], [64]).

При применении нейронных сетей следует учитывать присущий им параллелизм, т.е. использовать при этом параллельные компьютеры, имеющее подходящую структуру. Например, можно использовать мульти-процессорные системы, транспьютерные системы и т.п. Существуют реализации нейронных сетей в виде цифровых и аналоговых микросхем, а также реализации ANN с помощью нелинейных оптических систем.

2.1. Информационная структура эксперимента в ФВЭ

Для понимания последующего материала о применении ANN в ФВЭ представляется целесообразным дать основное представление об экспериментах в этой области и их информационной структуре.

Одной из особенностей экспериментальных исследований в физике высоких энергий является необходимость получения и обработки большого количества информации за короткое время. Это связано с принципиальной недетерминированностью процессов взаимодействия в микромире. При этом точность определения различных физических параметров увеличивается пропорционально корню квадратному из числа зарегистрированных событий изучаемой реакции. Зачастую доля изучаемых реакций составляет небольшую часть (10^{-6} и меньше) от всех возможных случаев реакций, которые тем не менее необходимо зарегистрировать. Типичные объёмы информации составляют $\sim 10^9 - 10^{11}$ событий/эксперимент в зависимости от типа эксперимента.

В физике высоких энергий существуют два основных типа ускорительных экспериментов: *эксперименты с фиксированной мишенью и на встречных пучках*.

При экспериментах с фиксированной мишенью выведенный из ускорителя пучок частиц, содержащий либо непосредственно выведенные из ускорителя протоны, либо смесь вторичных адронов (пионов, каонов, протонов или антипротонов — π , K , p , \bar{p}), полученных при взаимодействии первичного протонного пучка с внутренней мишенью ускорителя, направляется на покоящуюся *мишень установки* (отсюда и название). Происходящие в результате взаимодействия пучка с веществом мишени реакции регистрируются установкой.

Если реакция происходит при столкновении двух ускоренных пучков частиц (обычно это столкновения протонов-антипротонов $p\bar{p}$ или электронов-позитронов e^+e^-), направленных навстречу друг другу, то это — эксперимент *на встречных пучках*.

Обычно для экспериментов на выведенном пучке создаётся универсальная установка для изучения широкого класса реакций взаимодействия при высоких энергиях, например [10,11]. Состав оборудования, как правило, включает в себя сцинтилляционные и черенковские счётчики, пропорциональные и дрейфовые камеры, сцинтилляционные годоскопы и счётчики полного поглощения (калориметры), быстродействующую электронику для предварительного отбора событий, систему сбора данных и несколько компьютеров с аппаратурой стыковки с электроникой установки, средствами визуализации и периферией для записи данных.

Современный физический эксперимент на ускорителях настоящего и последующих поколений (У-70, SPS, Tevatron, LEP, HERA, УНК, LHC, RHIC) [34] характеризуется интервалом времени между столкновениями 15 – 100 нс. Вероятность взаимодействия зависит от сечения реакции и от *интенсивности* пучка или *светимости* ускорителя. Таким образом, информационные потоки в экспериментах определяются сечениями исследуемых процессов и возможностями регистрирующей аппаратуры. Их, в принципе, можно регулировать изменением интенсивности первичного пучка и толщины мишени.

К примеру, в мишени установки [10] за один импульс ускорителя происходило около 10^6 случаев различных реакций взаимодействия частиц пучка с веществом мишени. Из них примерно 500 событий относились к исследуемому в эксперименте классу реакций, причём только в $\sim 10\%$ случаев вторичные частицы попадали в телесный угол, перекрываемый детекторами установки, т.е. могли быть зарегистрированы.

Регистрирующая аппаратура типичной установки состоит, как минимум, из двух подсистем. Одна из них, включающая в себя только “быстрые детекторы” (сцинтилляционные и черенковские счётчики), регистрирует все события взаимодействия и передаёт информацию о каждом из них логической электронике предварительного отбора событий (выработки триггера), которая должна за ~ 50 нс обработать эту информацию и решить (дать триггер), принадлежит ли данное событие к классу изучаемых и надо ли включать вторую подсистему, чтобы полностью зарегистрировать выбранное событие. Вторая подсистема состоит из отно-

сительно “медленных устройств”: трековых детекторов, калориметров и модулей преобразующей и запоминающей электроники.

Во время выработки решения о регистрации данного события ($\sim 150 - 350$ нс) информация об этом событии запоминается в виде ионизированных следов заряженных частиц в камерах и в виде электрических импульсов, распространяющихся по длинным соединительным кабелям от детекторов к электронике. Объём исходной информации о полном зарегистрированном событии определяется возможностями системы сбора данных, а частота таких событий задаётся быстродействием этой аппаратуры.

Предварительный отбор событий (триггер) сокращает объём регистрируемой информации на несколько порядков и позволяет исследовать сравнительно редкие процессы, а также использовать относительно медленные трековые и калориметрические детекторы, медленные преобразователи и регистраторы в быстродействующей установке.

В зависимости от сложности задачи создаётся *триггерная система*, состоящая из триггеров различных уровней. Триггер нулевого уровня (иногда отсчёт ведётся от 1) служит для выделения пучковой частицы и акта взаимодействия в мишени, он же определяет *временную привязку*. Время выработки триггера $10 - 100$ нс. Скорость потока событий на входе $\simeq 10^7$ s^{-1} , на выходе — примерно на два порядка ниже.

Выделение определенного класса реакций (например, реакций с конечным нейтральным состоянием или с двумя заряженными треками) выполняется триггером первого уровня и занимает $\simeq 1$ мкс, скорость поступления событий “сбивается” ещё примерно на порядок.

И, наконец, триггер второго уровня вычисляет уже некоторые параметры реакции (массу продуктов реакции, определение типа частиц и т.д.), необходимое время $\simeq 10$ мкс.

Во время выработки триггера и чтения данных система сбора данных блокирует поступление сигналов о взаимодействиях. Длительность блокировки задаёт *мертвое время* установки. Очевидно, что для повышения эффективности работы скорость выработки триггера должна быть максимальной.

Вся информация о зарегистрированном событии передаётся из системы сбора данных в один или в несколько компьютеров, которые обрабатывают эту информацию в реальном масштабе времени проведения эксперимента с целью контроля работы установки и хода эксперимента в целом (on-line обработка). Затем эта информация записывается на магнитные или магнито-оптические носители долговременного хранения для последующей полной и, возможно, неоднократной обработки с целью получения окончательных физических результатов (off-line обработка).

Итак, в результате взаимодействия образуются и распадаются частицы, и задача регистрации реакции сводится, в конечном счете, к выделению акта взаимодействия определённого класса и восстановлению 4-импульсов продуктов реакции.

Для восстановления 4-импульса (\mathbf{p}, E) какой-либо частицы необходимо измерить пространственные координаты её следа (трека) и её энергию. Пространственные

координаты определяются *трековой системой*, в состав которой обычно входят:

- *Многопроволочные пропорциональные камеры. Принцип действия основан на возникновении электронной лавины при прохождении заряженной частицы через рабочее вещество камеры, её усилении и дрейфе к сигнальной проволоке в электрическом поле. Пространственное разрешение $\sigma \simeq 1$ мм.*
- *Дрейфовые камеры. Принцип работы имеет много общего с пропорциональной камерой, но дополнительно измеряется время дрейфа электронного облака к сигнальной проволоке. Пространственное разрешение $\sigma \simeq 0,1$ мм. В отличие от пропорциональной, дрейфовая камера имеет меньшее быстродействие, так как она имеет “память”, определяемую временем дрейфа.*
- *Времяпроекционные камеры. Усовершенствованный вариант дрейфовой камеры — дополнительно прикладывается магнитное поле и увеличен дрейфовый промежуток. Позволяет получить истинный пространственный трек частицы.*

Измерение энергии частиц производится *калориметрами*. В этих детекторах энергия частицы полностью поглощается и преобразуется в излучение другой природы (ионизационное, излучение Вавилова-Черенкова). Рабочим веществом при регистрации фотонов и электронов служит свинцовое стекло, различные неорганические кристаллы, а при регистрации адронов — чередующиеся слои сцинтиллятора и тяжелого вещества. Обычно калориметры имеют ячеистую структуру, характерные размеры ячейки — сантиметры. *Пространственное разрешение* — несколько миллиметров, *энергетическое* — проценты (зависит от энергии частицы).

2.2. Применение в триггерах

Триггеры в ФВЭ предназначены для многоуровневой фильтрации событий. Они должны работать быстро и точно. Особенно жёсткие требования предъявляются к системам регистрации и предварительного отбора событий (триггерам) в экспериментах на встречных пучках.

В современных коллайдерах интервал между событиями очень мал. Если на Теватроне он составляет около 400 нс, то на HERA — уже 96 нс, а на LHC он будет понижен до 16 нс. Кроме того, в последнее время наблюдается значительный рост сложности экспериментов в ФВЭ. Объём информации, снимаемой с детекторов, постоянно увеличивается.

ANN была применена в триггере второго уровня в H1 эксперименте (см. работу [53]). Задача этого триггера состояла в понижении частоты событий на HERA e-p коллайдере с около 0,5 МГц до нескольких герц. Эта огромная частота обуславливается фоном, идущим от сильных взаимодействий фотонного пучка. Частота “полезных” электрослабых физических процессов намного ниже. Для событий выше $Q^2 > 10$ ГэВ ожидаемая частота около 3 Гц. Фильтрация событий происходит на четырех уровнях: двух аппаратных уровнях (hardware) (L1 и L2), уменьшающих частоту до 100 Гц, и двух программных уровнях (software) (L3 и L4), дающих уменьшение до 5 Гц.

L1 получает информацию из каллориметра, центральных трековых камер и от время-пролётных (time-of-flight) детекторов. Большое количество этой информации может использоваться только частично, так как существует временное ограничение для L1 (2 мс). Для L2 временное ограничение также довольно сильное (10 мс). Для обработки информации на этом уровне была использована FFNN с единственным выходным нейроном. Сеть использовала топологические энергетические суммы и информацию, получаемую с триггера первого уровня. Тренировка осуществлялась с помощью пакета Jetnet 2.0 на образцах физических событий и фона, смоделированных методом Монте-Карло. Вектор образца насчитывал 19 элементов. Шум и события были представлены в одинаковой пропорции.

Для классов физических событий максимальное отделение было достигнуто с 25 промежуточными нейронами. Использование сети с 4 слоями в сравнении с тремя не улучшило эффективность работы сети. Для тренировки сети достаточно от 50 000 образцов до нескольких миллионов в зависимости от структуры сети.

Исследования показали, что L2 триггер с ANN сдерживает фон на приемлемом уровне. Для срабатывания сети за 10 мс требуется вычислительное устройство с производительностью обработки около 120 миллионов взаимодействий/секунду. Данный алгоритм реализован благодаря использованию специальных быстрых микросхем (matrix-vector multiplier chips).

Для распознавания треков в мюонной камере FNAL была использована аналоговая ANN, которая была построена на специальных аналоговых микросхемах (Intel's ETANN). События имели импульс больше чем 5 ГэВ/с. В работе [4] можно найти описание применения ANN в триггере мюонных треков. Нейронная сеть позволила повысить точность в определении импульса и сократить время вычислений. Успех использования сети заключается в том, что были использованы аналоговые времена дрейфа (непосредственно, без их оцифровки), что обеспечило хорошую скорость и точность. FFNN имела 12 входных, 64 промежуточных и 64 выходных нейронов. 32 выходных нейрона определяли наклон трека, а другие 32 — его место расположение (пересечения). Тренировка сети была проведена в режиме off-line с помощью *BP* на 10.000 треках, смоделированных методом Монте-Карло.

Работа сети на реальных треках дала разрешение 1,2 см, что только в два раза хуже, чем разрешение, полученное в off-line анализе, и в 4 раза лучше, чем полученное с помощью обычных методов, применяемых в триггерах. Время, требуемое для получения выходного результата для триггера, составляет ~ 8 мкс.

В эксперименте E-735 FNAL ($p - \bar{p}$ столкновение) с помощью FFNN в триггере второго уровня решается задача быстрой реконструкции треков в дрейфовой камере (см. работу [4]). Реализация основана на Intel ETANN микросхемах. Сеть находит первичные вершины взаимодействия при столкновении пучковых частиц. Поперечная координата вершины хорошо известна благодаря ограниченному размеру пучков сталкивающихся частиц, а продольная координата z является неопределенной. В работе были использованы 18 перекрывающихся сетей. Входными данными для сетей были времена дрейфа. Каждая сеть имела 18 входных, 128 промежуточных и 62 выходных нейронов. Выходом каждой сети является z координата первоначаль-

ного столкновения, представленная в виде гистограммы (определяющей плотность вероятности z координаты события). Для достижения большей точности выходные гистограммы складываются. Полученная координата вершины согласуется с более точной off-line реконструкцией и дает пространственное решение в три раза лучше, чем другие on-line методы, основанные на время-пролётной информации.

ANN входит в блок электроники, основанного на ETANN микросхемах, в изолирующем (isolation) триггере 2-го уровня в CDF ($p-\bar{p}$ -столкновение в эксперименте FNAL) (см. работу [4]). Наличие изолированного электрона указывает на распад W -бозона. Для идентификации электронного ливня была использована FFNN, имеющая 25 входных, 4 промежуточных и 1 выходной нейроны. Сеть получает на вход массив калориметрических данных размером 5×5 . Электронный ливень может располагаться в этом массиве четырьмя способами. Каждый способ представлен как промежуточный нейрон в сети. Сеть распознает структуру, аналогичную одному из этих четырех способов. Она вычисляет энергию внутри и вне предполагаемого ливня и сообщает о наличии изолированного электрона, если, по крайней мере, один промежуточный нейрон имеет $f \times E_{inner} - E_{outer} > 0$. Путем тренировки сети найдена наилучшая величина f равная 0.16. Применение нейронной сети в блоке электроники ускорило обнаружение изолированного электрона. Увеличение скорости данной операции позволило переместить ее с третьего уровня триггера на второй. Благодаря использованию сети удалось понизить энергетический порог для электронов и сократить фон в 4 раза.

2.2.1. Триггер на b -кварк

События, содержащие тяжёлые кварки (конкретно b) очень редки по сравнению с фоновыми событиями, и отношение сигнал/фон для них составляет от 10^{-3} (на адронных коллайдерах) до $\simeq 10^{-5}$ (эксперименты с фиксированной мишенью). Улучшить это соотношение возможно при использовании триггера, который отбирал бы события с характерными для тяжелых кварков признаками. Как уже упоминалось, это может быть, например, лептон с большим поперечным импульсом от распада B -мезона. Недостатком является то, что возможны и другие процессы с испусканием лептона, и кроме того, только часть b -кварков распадается по этой моде. Идеальным решением проблемы был бы триггер на события [16], содержащие треки с началом не в точке взаимодействия (вторичные вершины).

На установке WA92 в CERNe используется микровершинный детектор, состоящий из системы микростриповых камер. Сигналы оцифровываются и подаются на вход ассоциативной памяти, которая выдает список всех допустимых треков для данного набора хитов. Треки параметризованы двумя переменными — импакт-параметром d и углом Φ . Таким образом, задача выделения событий с b -кварками сводится к тому, чтобы по набору (D, Φ) определить, есть ли в событии вторичные вершины.

Для решения задачи была использована FFNN, реализованная на СБИС ETANN 80170NX, со структурой 64–64–8. На вход подавались величины D и Φ для каждой

плоскости. События с b -кварком первые четыре выхода должны были устанавливаться в 1, а остальные — в -1 , а фоновые события соответственно наоборот. С целью обучения было сгенерировано 2500 фоновых событий и 500 событий с b -кварком. После обучения подавление фона составило 99% при эффективности регистрации полезных событий 30%. Ранее такое качество обогащения b -событиями было возможно только при off-line анализе. Весь процесс выработки триггера занимал несколько микросекунд.

2.3. Классификация и идентификация событий

Одной из наиболее типичных задач ФВЭ для ANN является выбор и классификация событий. Как правило, события характеризуются больше, чем одной переменной, граница между ними часто оказывается сильно нелинейной, и выбор оптимального сечения не очевиден. На помощь приходит MLP. Функцию перехода для таких сетей выбирают гладкой, обычно берут функцию сигмоидного типа (см. формулу (3)). Такой выбор гарантирует, что функция ошибки будет изменяться гладко при тренировке сети с помощью BP , который требует существования производной у функции перехода. Выбор функции сигмоидного типа обуславливается еще и тем, что для перекрывающихся классов она может быть использована для аппроксимации вероятности того, что событие относится к одному или другому классу.

При использовании BP для получения наилучшего результата желательно проводить предварительную обработку данных с целью выбора “разумных” переменных, так как для того, чтобы сеть могла “охватить” все присущие “сырым” данным свойства, потребуется очень большой тренирующий ряд. Часто для тренировки сети берутся события, смоделированные методом Монте-Карло. В этом случае необходимо учитывать тот факт, что работа сети зависит от выбранной модели.

К задачам классификации, успешно решаемым с помощью ANN, относятся:

- *разделение электромагнитных и адронных ливней в калориметрах,*
- *разделение кварк-глюонных струй,*
- *идентификация частиц, в том числе, непосредственно не наблюдаемых (кварков), распадов, систем частиц в конечном состоянии,*
- *идентификация частиц по ионизации, по кольцам черенковского света и т.п.*

2.3.1. Разделение кварк-глюонных струй

В ФВЭ струя определяется как система движущихся частиц с малым угловым распределением и ограниченным переданным импульсом по отношению к оси струи. Анализ струй, возникающих в процессах e^+e^- -столкновений и глубоконеупругого соударения, позволяет глубже изучить процессы адронизации кварков и глюонов, прояснить механизм конфайнмента (невозможности наблюдения кварков в свободном состоянии).

Безмодельное экспериментальное сравнение струй от кварков и глюонов, которое можно сделать для некоторых классов реакций [31], а также расчёты КХД показывают, что глюон рождает в среднем более “рыхлую” струю, с меньшей энергией в центральном конусе, и большей множественностью частиц, чем кварк.

Существуют следующие способы представления струи с помощью кинематических переменных [36,37]:

- 4-импульс (\vec{p}_k, E_k) — 4-х лидирующих частиц в струе,
- 4-импульс струи (p_{jet}, E_{jet}) и лидирующей частицы (\vec{p}_p, E_p) .

Тем не менее перекрытие распределений этих кинематических переменных струи велико, и простой отбор по ним не приводит к достоверному восстановлению источника. Таким образом, идентификацию источника струи можно рассматривать как задачу распознавания образов, для решения которой использовать ANN, и прямой способ идентификации струи состоял бы в нахождении функционального отображения набора существенных кинематических параметров струи и типа источника струи — кварка или глюона.

Струя, образованная кварком (глюоном), регистрируется трековой системой и калориметрами с периодической ячеистой структурой. Для идентификации струи используется традиционная feed-forward NN — FFNN с архитектурой 16–10–1 или 6–6–1 (зависит от типа представления струи). Данные, полученные с помощью программ Монте-Карло, имитирующих различные теоретические модели (JETSET 7.2, ARIADNE 3.1 и т.д.), подаются на вход первого слоя сети. Обучение происходит с помощью метода обратного распространения ошибок (Back Propagation Error, в дальнейшем *BPE*). Соотношение числа событий для обучения/тестирования = 2/1. После $\simeq 10^3$ эпох обучения получена эффективность сепарации струй до 85%, что близко к теоретическому пределу. С помощью простых обрезаний по различным кинематическим переменным удается достичь 65% идентификации источника струи.

Позднее, в работах [12], [9] был развит так называемый k_{\perp} алгоритм, который имеет хорошее теоретическое обоснование. Для определения принадлежности частицы i к струе j в e^+e^- -столкновениях вводится переменная

$$d_{ij}^{k_{\perp}} = 2 \min(E_i^2, E_j^2) (1 - \cos\theta_{ij}), \quad (27)$$

и величина $y_{ij} = d_{ij}^{k_{\perp}}/Q^2$ (Q — суммарная энергия струи) сравнивается с некоторым пороговым значением y_{cut} (характерное значение для принадлежности к струе $\simeq 10^{-2} - 10^{-1}$). Для адрон-адронных столкновений алгоритм приписывает частицу i или струе j или считает ее принадлежащей пучку в зависимости от разницы между

$$d_{ij} = 2 \min(E_{Ti}^2, E_{Tj}^2) \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \quad (28)$$

и

$$d_{iB} = E_{Ti}^2, \quad (29)$$

где E_{Ti} — переданная энергия частицы i в проекции на направление пучка, η_i — ее псевдобыстрота, ϕ_i — азимутальный угол по отношению к оси пучка.

k_{\perp} алгоритм достаточно сложен для прямых вычислений, и в работе [15] была поставлена задача симулировать работу алгоритма с помощью ANN.

События генерировались программой HERWIG [39], энергия пучка $\sqrt{s} = 1,8$ ТэВ. Принимались во внимание все частицы, образованные при генерации и принадлежащие струе ($n_p \leq 80$). Использовались две различные архитектуры сети — традиционная FFNN и самоорганизующая сеть (обучение “без учителя”).

Целью было определение количества струй в событии, сравнение качества работы сети и прямых вычислений по алгоритму k_{\perp} . Для FFNN была выбрана структура $240 - 100 - 5(7)$. Количество выходных нейронов определлось пределами классификации событий — до $5(7)$ струй в событии. Струя представлялась набором импульсов p_x, p_y, p_z всех частиц (≤ 80) в струе, отнормированных на интервал $[0,1]$ и упорядоченных по энергии. Если по работе k_{\perp} алгоритма событие относилось к классу с i струями, то при обучении требуемый выходы сети $Y_i = 1$ и $Y_k = 0, k \neq i$. Набор для обучения состоял из $\simeq 40 \cdot 10^3$ событий. Эффективность и чистота выделения событий регулировались параметром T_h , который определяется следующим образом: при тестировании работы сети выход

$$Y_i = 1 \text{ если } Y_i \geq 1 - T_h, \quad (30)$$

$$Y_i = 0 \text{ если } Y_i < 1 - T_h. \quad (31)$$

При $y_{cut} = 0,1$ события с 1 и 2 струями выделялись с эффективностью и чистотой 0,6; 0,98 и 0,6; 0,91 соответственно.

Отдельно была изучена возможность работы самоорганизующейся сети SOM^2 . Для этого была использована сеть с входным слоем из 240 нейронами. Выходной слой состоял из M нейронов, $5^2 < M < 20^2$ (лучшие результаты были получены с большим числом нейронов). На вход сети предъявлялось событие, и расстояние

$$d_k = |x - \mathbf{W}_k| = \sqrt{\sum_{i=1}^N (x_i - W_{ik})^2} \quad (32)$$

вычислялось для каждого выходного нейрона, W_{ik} — элемент синаптической матрицы, соединяющий входной нейрон i и выходной нейрон k . Первоначально элементам матрицы были присвоены малые случайные значения. Если выходной нейрон m имеет наименьшее расстояние

$$d_m \leq d_k \quad \forall k = 1, \dots, M, \quad (33)$$

то он объявлялся “победителем”, и только его вес модифицировался

$$W_{ij} \rightarrow W_{ij} + \delta W_{ij}, \quad (34)$$

²Self-Organizing Map.

$$\delta W_{ij} = \eta_j(x_i - W_{ij}). \quad (35)$$

Результатом такой операции будет приближение W_{ij} к x_i . “Скорость движения” $\eta_j = \eta \Lambda(j, m)$, где функция $\Lambda(j, m)$ максимальна при $j = m$ и резко спадает при увеличении $|j - m|$. Это гарантия того, что не только выходной нейрон-победитель, но и его ближайшие соседи изменят свои веса ближе к \mathbf{x} . В результате изменения весов по правилу 34, 35 и 34 матрица \mathbf{W}_m будет давать представление всех событий, которые делают нейрон m победителем.

Топология событий, имеющих одинаковое число струй, может сильно отличаться друг от друга, поэтому с практической точки зрения удобно иметь выходной слой с большим количеством нейронов. После самоорганизации сети на той же выборке событий (см. выше) можно выбрать два способа анализа синаптической матрицы W_{ij} .

События сначала обрабатываются k_{\perp} алгоритмом. Каждое событие после этого имеет “метку” n_j — множественность струй в событии. Определение чистоты и эффективности для каждого класса с l струями вводится следующим образом:

$$p_l = \frac{N_l^a}{N_{tot,l}^a}, \quad (36)$$

$$\eta_l = \frac{N_l^a}{N_l}, \quad (37)$$

$$N_l^a = \sum_{m|l} \omega_l^{(m)}, \quad (38)$$

где символ $m|l$ означает что в суммировании принимают участие все нейроны, представляющие класс с l струями.

$$N_{tot,l}^a = \sum_{m|l} \sum_j \omega_j^{(m)} \quad (39)$$

есть общее количество принятых событий. Для каждого класса событий селективность сети в терминах (чистота, эффективность) может быть получена модифицированным правилом 33:

$$d_m \leq \min\{t, d_k\} \quad \forall k = 1, \dots, M, \quad (40)$$

где t — внутренний параметр. Результаты показывают, что *SOM* хорошо работает для событий с малой множественностью струй (≤ 3).

Другой подход к *SOM* заключается в том, что \mathbf{W}_k рассматривается как внутреннее представление набора событий, возбуждающих нейрон k , а $\mathbf{W}_k = \{W_{ik}\}$ — как распределение импульсов частиц гипотетического W_{ik} события. Сеть используется как *модель* набора событий, который был использован при образовании W_{ik} и может анализироваться k_{\perp} алгоритмом. Число событий W_{ik} $M \leq 20^2$, что намного меньше первоначально использованного набора ($\simeq 40 \cdot 10^3$). Очевидно, что требуемое на анализ время будет в этом случае заметно меньше. Проверка показала, что работа k_{\perp} алгоритма на гипотетических W_{ik} событиях приводит к результатам, аналогичным при обработке всей выборки событий.

2.3.2. Идентификация тяжелых кварков

Определение источника струи — кварк или глюон — не всегда бывает достаточно. Развитием изложенного выше является определение типа кварка, который образовал струю.

Струя, образованная b -кварком, имеет следующие характерные признаки [40]:

- присутствие энергичной лидирующей частицы B (масса $\simeq 5$ ГэВ);
- доминирующая каскадная мода распада $B \rightarrow C + X, C \rightarrow S + X$, где C и S — адроны, содержащие c - и s -кварки; это приводит к тому, что и множественность $\langle n \rangle$ для струи от b -кварка больше, чем для обычной, а также специфическим корреляциям для продуктов распада b -кварка;
- заметный $BR(B \rightarrow e/\mu + X) \simeq 20\%$ с большими отношением энергий e/μ и p_T ;
- большое среднее время жизни частицы B ($\tau \simeq 10^{-12}$ сек.) приводит к наблюдаемому разрыву между точкой взаимодействия в мишени и точкой распада.

Процессы с образованием b -кварка являются достаточно редкими для процессов e^+e^- и $p\bar{p}$ соударений. Поэтому задачу выделения событий с b -кварком необходимо поставить на уровне триггера, или использовать уже в off-line анализе. Входными переменными для ANN [64] являются параметры, описывающие форму струи в калориметре (число ячеек калориметра с энергией больше пороговой E_{cell} , энерговыделение в части струи, выделяемой в круге радиуса R , 4-импульсы лидирующих частиц в струе). В зависимости от числа входных переменных (17–10–1, 20–10–1) структура сети менялась незначительно, и заметного влияния на конечные результаты не было обнаружено.

Результаты показывают, что качество выделения b -струи не уступает другим подходам — с использованием вершинного детектора или с учетом лептонной моды распада b -кварка. На установке CDF с аналогичным набором входных параметров для сети удалось получить 95%-ое подавление фоновых событий при 65%-ой эффективности регистрации струи.

2.3.3. Идентификация распадов τ -лептона

Выделение редких распадов элементарных частиц связано со значительными трудностями, которые возникают из-за тяжелых фоновых условий регистрации. Другие моды распада одной и той же частицы могут имитировать редкий процесс. Поэтому обычные методы выделения, основанные только на кинематическом фите, сталкиваются со значительными трудностями.

В частности, при выделении распадов τ -лептона на установке L3³ [29], [32] трудно разделить адронные моды распада

$$\tau \rightarrow \pi\nu, \tag{41}$$

³Детектор L3 состоит из центральной трековой системы, электромагнитного и адронного калориметров и из системы прецизионных мюонных камер. Вся установка находится внутри 12 м магнита с постоянным полем 0,5 Тесла.

$$\tau \rightarrow \rho\nu \rightarrow \pi\pi^0\nu, \quad (42)$$

$$\tau \rightarrow a_1\nu \rightarrow \pi\pi^0\pi^0\nu, \quad (43)$$

$$\tau \rightarrow \pi\pi^0\pi^0\pi^0\nu \quad (44)$$

из-за малого угла разлета продуктов распада τ ($\pi^-, \pi^+, \pi^0 \rightarrow \gamma\gamma$) — пространственного разрешения отдельных частей установки часто бывает недостаточно для разделения треков или ливней. Поэтому применение ANN с целью выделить внутренние корреляции в событиях, которые присущи продуктам распада τ -лептона, кажется довольно привлекательным.

Нейронная сеть была применена для идентификации распада $\tau \rightarrow \rho\nu$, наиболее чувствительного к поляризации τ -лептона, а также для $\tau \rightarrow \pi(K)\nu$ и $\tau \rightarrow e\nu\bar{\nu}$. Источником τ служит распад

$$Z \rightarrow \tau^+\tau^-, \quad (45)$$

выделяемый с незначительным фоном. Для представления информации о событии был выбран ряд измеряемых в эксперименте величин, которые описывают характер энерговыделения в калориметрах и трековую информацию.

Использовалась FFNN со структурой 18–18–1. Тренировка сети осуществлялась традиционным *BPE* алгоритмом на данных, полученных методом Монте-Карло с учетом симуляции отклика детекторов. Всего было сгенерировано $40 \cdot 10^3$ событий реакции $e^+e^- \rightarrow \tau^+\tau^-$. Цикл обучения состоял из одного сигнального события и одного фонового, веса обновлялись через каждые пять циклов. Для обучения требовалось $\simeq 3 \cdot 10^6$ циклов. Эффективность регистрации распада $\tau \rightarrow \rho\nu$ составила 54% с чистотой 83%. Проведенные тестовые испытания с внесением дополнительных искажений в отклик детекторов показали хорошую устойчивость сети при распознавании распадов.

2.3.4. Поиск хиггсовского бозона H

В последнее время появился ряд работ [6], [28], [57], где рассматривается применение ANN для поиска гипотетической частицы современной теории электрослабых взаимодействий - хиггсовского бозона H , ответственного за механизм нарушения симметрии. Задача поиска и обнаружения H является в настоящее время наиболее важной для ФВЭ. Трудности усиливаются еще и тем обстоятельством, что не существует достаточно точной теоретической оценки для массы H , и обычно она рассматривается в интервале 60 ГэВ — 1 ТэВ ⁴

В pp -взаимодействиях основным механизмом образования H является глюонное gg - или W^+W^- -смешивание. При распаде по каналу

$$H \rightarrow Z^0 Z^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^- \quad (46)$$

⁴По последним данным $m_H \leq 650$ ГэВ.

физически неустранимым источником фона служит образование пар Z -бозонов КХД континуумом с последующим распадом по каналу $\mu^+\mu^-$. Расчеты с помощью программы RUTHIA показывают, что для тяжелого H с массой $m_H = 750$ ГэВ отношение сигнал/фон изменяется от $\simeq 10^{-3}$ до $\simeq 10^{-1}$ в зависимости от энергии в системе центра масс (от $\sqrt{s} = 16$ до $\sqrt{s} = 200$ ТэВ соответственно). Для разделения событий этих двух классов была использована FFNN со структурой 8–5–1 и входными параметрами, сконструированными по следующему алгоритму:

- 1) отобрать все частицы, попавшие в конус с апертурой $0,8$ в пространстве $\eta - \phi$ (η — псевдобыстрота, ϕ — азимутальный угол);
- 2) выделить струю с максимальным импульсом;
- 3) выделить 4 самых энергичных частицы, принадлежащих к лидирующей струе;
- 4) вычислить входные переменные для ANN, $|p|$, $|p_t|$ (относительно оси струи).

Для дальнейшего обучения важно проверить, какие частицы являются лидирующими в струях, образованных от распада H , и от фоновых событий. Оказывается, что в случае событий от H лидирующей частицей наиболее часто оказывается μ (41%), а для фоновых — π и μ (соответственно 29 и 26%). γ -квант бывает наиболее энергичной частицей в струе в 21% для фоновых и 16% для H -событий. Различие в сортах лидирующих частиц приводит к различным корреляциям в переменных $|p| - |p_t|$, и на этом различии и предлагается выделить события образования H . Проведенные расчеты показали, что после тренировки FFNN позволяет улучшить соотношение сигнал/фон более чем на порядок. Это превосходит по эффективности другие обычно применяемые способы.

Тщательный анализ применения ANN был выполнен в работе [57] для поиска H с массой 70 (90) ГэВ в процессах e^+e^- -столкновений

$$e^+e^- \rightarrow HZ \quad (47)$$

на будущем ускорителе LEP-200 с последующим распадом $H \rightarrow q\bar{q}$ и $Z \rightarrow q\bar{q}$. Сечение фоновых процессов

$$e^+e^- \rightarrow ZZ, \quad (48)$$

$$e^+e^- \rightarrow W^+W^-, \quad (49)$$

$$e^+e^- \rightarrow q\bar{q}gg, \quad (50)$$

как и в случае pp -столкновений, превосходит на два порядка величины процесс образования Хиггса. Стандартный набор обрезаний состоит из отбора событий с 4 струями в конечном состоянии, удаления событий с энергичным ($E_\gamma > 30$ ГэВ) γ -квантом, часто испускаемым в процессе, а также с учетом того обстоятельства, что струя от H обогащена b кварками, и следовательно, имеет большую множественность частиц со вторичной точкой распада.

Для выделения H была использована FFNN с архитектурой 45–22–1. Анализ чувствительности сети по различным входным переменным позволил без потери общности уменьшить их число до 10:

1. N_{off} — число вторичных треков.
2. χ^2 , когда событие фитируется гипотезой HZ .
3. $E_{\gamma max}$ в событии.
4. N_{ch} — число заряженных треков.
5. T — показатель анизотропии события.
6. χ^2_{WW} при фитировании события WW -гипотезой.
7. m_{off} — инвариантная масса всех вторичных треков.
8. S_{θ} — сумма углов между струями для событий с 3 струями.
9. P_{max}^e — импульс самого энергичного электрона.
10. N_{off}^H — число вторичных треков в H -струе при фитировании события гипотезой HZ .

Для обучения сгенерированы $8 \cdot 10^4$ ZZ -событий, $2,4 \cdot 10^5$ WW -событий, $8 \cdot 10^5$ $q\bar{q}$ -событий и $5 \cdot 10^4$ сигнальных событий. По сравнению со стандартными отборами ANN показывает на $\simeq 60\%$ большую эффективность в терминах светимости, т.е. для получения статистической значимости в 5 стандартных отклонений надо иметь на 60% событий меньше.

2.3.5. Идентификация системы частиц в конечном состоянии

Эксперимент CPLEAR изучает эффекты нарушения чётности С/СР/Т в $K^0 - \bar{K}^0$ системе [65]. Детектор CPLEAR имеет цилиндрическую структуру. иждководородную мишень окружают два слоя проволочных пропорциональных камер (РС), шесть слоёв дрейфовых камер (DC), два слоя стримерных трубок (ST), сцинтиляционные (S1, S2) и черенковские (С) счетчики. Вся трековую систему окружает газовый калориметр. Установка находится в магнитном поле напряженностью 0,44 Тесла. Система выработки триггера состоит из трех уровней и выполняет следующие задачи:

- Идентификация событий, содержащих возможные K^{\pm} треки и, по меньшей мере, одну заряженную частицу. Решение вырабатывается за время < 60 нс.
- Идентификация сорта частиц (по времени пролета, удельным ионизационным потерям). Решение вырабатывается за $\simeq 0,5$ мкс.
- Вычисление переданного импульса, фитирование и параметризация треков в событии. Требуется $> 0,5$ мкс.

Нейтральные каоны образуются в реакции аннигиляции

$$p\bar{p} \rightarrow \pi^{\pm} K^{\mp} K^0(\bar{K}^0) \quad (51)$$

с вероятностью $\simeq 0,2\%$. Ассоциированное образование K^0 с K^- и \bar{K}^0 с K^+ позволяет “метить” странность образующихся каонов. При использовании величин, доступных на второй стадии триггера, возможно применение двух классических способов отбора, основанных на кинематических переменных:

1. Отбор по недостающей массе M_{miss} . При этом принимается 60% событий, из которых 70% хороших.
2. Отбор по сумме импульсов двух первичных частиц (65 и 90%).

Существующая система отбора событий увеличивает долю событий реакции (51) в $\simeq 100$ раз, но при дальнейшем off-line анализе 60% событий исключаются. С целью улучшить качество триггера был предложен способ, основанный на обработке событий с 4 треками в конечном состоянии и вычислении их кинематических параметров.

Для выделения реакции (51) необходимо учитывать, что при аннигиляции интенсивно идут и другие реакции, которые можно разбить на следующие классы:

1. $p\bar{p} \rightarrow K^\pm \pi^\mp K^0, K^0 \rightarrow \pi^+ \pi^-$.
2. $p\bar{p} \rightarrow K^\pm \pi^\mp K^0, K^0 \rightarrow \pi e \nu$.
3. $p\bar{p} \rightarrow K^\pm \pi^\mp K^0, K^0 \rightarrow \pi^+ \pi^- \pi^0$.
4. $p\bar{p} \rightarrow \pi^0 K^\pm \pi^\mp K^0$.
5. $p\bar{p} \rightarrow K^+ K^- \pi^+ \pi^-$.
6. $p\bar{p} \rightarrow \pi^+ \pi^- \pi^+ \pi^-$.

События из классов 1–3 должны быть оставлены для дальнейшего анализа, так как отличаются только модой распада K -мезона, а события из классов 4–5 являются фоновыми и должны быть исключены.

Частицы, образующиеся непосредственно в $p\bar{p}$ -аннигиляции, назовем первичными, а продукты распада K -мезона — вторичными. Очевидно, что сами каоны всегда являются первичными частицами. Для каждого трека вычисляются 4 параметра:

- 1) импульс частицы в поперечной плоскости p_t ,
- 2) компонента импульса p_z ,
- 3) минимальное расстояние трека от центра детектора ϵ (в дальнейшем не рассматривается),
- 4) угол ϕ с осью x касательной в точке, ближайшей к центру детектора.

Входными параметрами для сети служат разности углов треков от первичных и вторичных частиц, а также величины p_t и p_z . Эти величины вычисляются в on-line. Импульс выражается в ГэВ/с, угол представляется его косинусом. p_t всегда больше 0, и удобнее использовать величину $2p_t - 1$. Таким образом, входные переменные частично нормализованы и лежат в интервале $(-1, 1)$.

ANN имеет обычную FF архитектуру с одним внутренним слоем (10 узлов). Были проведены работы по исследованию влияния архитектуры сети (количество входных переменных, внутренних слоёв и узлов), которые показали достаточность такого набора для удовлетворительной работы.

Обучение производится BP алгоритмом, функционал ошибки имеет вид

$$E = \sum_{\mu} \left(T^{\mu} \ln \left(\frac{T^{\mu}}{O^{\mu}} \right) + (1 - T^{\mu}) \ln \left(\frac{1 - T^{\mu}}{1 - O^{\mu}} \right) \right), \quad (52)$$

T^μ — предписываемый выход, O^μ — реальный выход сети.

Функционал стремится к бесконечности при увеличении разности между реальным и предписываемым выходом сети, и такое поведение естественно при большой ошибке.

События всех 6 классов генерировались программой Монте-Карло CPGEANT. Выборка для обучения состояла из 3000 событий для классов 1–3 и 4500 событий для классов 4–5. Температура промежуточного слоя была равна 5, а выходного — 2. Такой выбор разных температур диктовался необходимостью четкого разделения событий из разных классов.

Проверка работы на реальных данных показала, что поток событий, направляемый для off-line анализа, уменьшается вдвое, при этом отбрасывается 20% хороших событий (из классов 1–3).

Алгоритм был реализован на процессорах общего применения R3051 с RISC архитектурой. Тем не менее временной интервал, в течение которого должно быть выработано решение (25–50) мкс, слишком велик. Поэтому выбор был остановлен на коммерчески доступном сигнальном процессоре DSP96002, выполненном в стандарте VME. Это упростило связь с системой сбора данных установки CPLEAR. Время принятия решения составило 38 мкс.

Позднее, для улучшения качества триггера первого уровня, была создана и испытана электронная система, реализующая алгоритм работы нейронной сети, на основе быстрых стандартных микросхем ТТЛ [67]. Целью ANN триггера является определение числа треков заряженных частиц и их локализации в событии. Для идентификации трека используются все 11 слоев детектора CPLEAR. При считывании информации она разбивается на 64 угловых сектора. 5 соседних секторов обрабатываются одной платой, конструктивно выполненной в стандарте VME. Поэтому на вход ANN поступает 5×11 бит, внутренний слой состоит из двух узлов. Если выход сети установлен в логическую единицу, то треку приписывается значение среднего углового сектора.

Сложные арифметические операции трудно выполнить за ограниченное время (< 1 мкс), поэтому хранение весов и операции умножения/сложения реализованы табличным способом (Look Up Table — LUP) на быстрых SRAM⁵ микросхемах. Чтобы размер памяти имел разумные пределы, входы каждого внутреннего узла обслуживаются 4 SRAM. Выходы всех SRAM, относящихся к одному узлу, суммируются и ужимаются до 8 бит. Нелинейная часть активационной функции нейронов внутреннего и выходного слоев тоже хранится в SRAM. Все памяти загружаются весами через VME шину, что позволяет тестировать блоки и логику их работы.

При разработке было обнаружено, что специальные сумматоры с оптимизированной архитектурой, выполненные на FPGA⁶ не дают большого преимущества по скорости выполнения операций, но значительно дороже по стоимости. Время выдачи решения не зависит от конфигурации входных бит и составляет 60 нс, что достаточно для применения на ранней стадии выработки триггера.

⁵SRAM — Static Random Access Memory.

⁶Field Programmable Gate Arrays.

Поскольку объём памяти для табличных вычислений невелик, то была проверена погрешность, возникающая из-за ограниченной точности операций. Проверка показала расхождение в результатах на уровне 10^{-4} . Работа ANN сравнивалась с другими алгоритмами по следующим критериям:

- неэффективность восстановления трека,
- неправильное восстановление трека.

Результаты ANN практически не уступают off-line программе реконструкции, но значительно превосходят ее по быстродействию, а по качеству заведомо превосходят существовавший ранее триггер. Важно отметить, что структура сети в полной мере соответствует детектору и допускает перепрограммирование на другие задачи.

2.3.6. Идентификация частиц по энергетическим потерям

В эксперименте CPLEAR ANN используется для быстрого определения сорта частиц по энергетическим потерям dE/dx . При малом импульсе (< 500 МэВ/с) спектр энергетических потерь пионов и каонов различен, хотя и перекрывается. Идея быстрого определения сорта частиц заключается в применении сигналов, доступных на ранней стадии выработки триггера. Энергетические потери определяются по количеству света, выделенному в сцинтилляторе и измеренном фотоумножителем. Импульс частицы измеряется традиционным магнитным спектрометром. Переданный импульс p_T примерно может быть вычислен по разности номеров проволок между внутренней и внешней дрейфовыми камерами. Проекция импульса заряженной частицы на ось пучка может быть получена из p_T с учетом координаты z точки взаимодействия.

Таким образом, для ANN образуется следующий набор входных величин:

1. Сумма оцифрованных сигналов фотоумножителей, откалиброванных на единицу для сигнала от минимально ионизирующей частицы.
2. Разница сигналов фотоумножителей для информации о z -координате частицы, с разрешением $\simeq 20$ см.
3. Разница в номерах проволок внутренней и внешней дрейфовых камер для информации о переданном импульсе.

Архитектура сети 3–6–6–4. Обучение сети производилось на 2000 образцах, содержащих как пионы, так и каоны. Выход ANN для кандидатов в пионы был $(0.8, -0.8, -0.8, 0.8)$, для каонов — $(-0.8, 0.8, 0.8, -0.8)$. Использование четырех выходов вместо одного объясняется тем, что проведенные ранее исследования [17], [18] показали большую устойчивость и лучшую обучаемость для такой архитектуры ANN, чем в случае одного или двух выходов. Для сходимости при обучении, проведенном *BPE* методом, потребовалось 7000 эпох. Параметр обучения, определяющий скорость изменения весов, в процессе тренировки изменялся от 0,2 до $0,5 \cdot 10^{-2}$. Ошибка Δ , определяемая как

$$\Delta^2 = \frac{1}{n} \sum_{k=1}^n (O_k - T_k)^2, \quad (53)$$

где n — число выходов, O — действительный выход и T — предписываемый выход, в конце обучения составляла $\simeq 20\%$.

Проведенный тест показал, что 88% пионов и каонов правильно классифицируются, если считать, что выходы узлов 1 и 4 больше 0 и выходы узлов 2 и 3 меньше 0 для каонов и, соответственно, наоборот для пионов.

2.3.7. Идентификация частиц в детекторе типа RICH

Детектор колец черенковского излучения (Ring Imaging CHerenkov — RICH) предназначен для идентификации частиц (π, K, p) по размеру кольца черенковского излучения [1]. При прохождении частицы через радиатор счетчика генерируется конус излучения, которое детектируется фоточувствительным детектором, состоящим из 12×12 падов. Медленные частицы образуют узкое, тонкое и плотное кольцо, а пионы — более широкое и менее интенсивное. Таким образом, задача идентификации частиц сводится к известной проблеме распознавания образов. Традиционный способ заключается в выделении окружностей и фитировании для точного определения радиуса.

В процессе разработки были испытаны несколько вариантов архитектуры нейронной сети, построенной на СБИС 80170NX. Входы ANN разбиваются на две группы: внешние и входы обратной связи. Каждая группа содержит 80×64 синаптических весов (включая 16 постоянных смещений). Входы обратной связи могут быть использованы для отображения 128 входов в 64 выхода, используя внутренний мультиплексор. Во всех случаях сеть состоит из трех слоев. Каждый слой, в свою очередь, может быть разбит на модули, равные одной СБИС, которых в слое может быть несколько.

Настоящая система обучалась с использованием до восьми СБИС. Максимальное число входов при таком способе достигает 8×128 , а выходов — 8×64 . Сигнал проходит слой за $\simeq 1 - 3$ мкс. Для обучения используется *VPE* алгоритм, требующий 300 эпох для набора из 700 пар-образцов. В конце обучения средняя ошибка понижается до $2 \cdot 10^{-3}$. На 98% представленных образцов сеть достоверно отделяла протон от π -мезона. При работе фактически были использованы две платы VME, соединенные 96-контактным кабелем. Одна содержала собственно СБИС ETANN, а другая служила картой управления и включала двухпортовую память емкостью 2 Мбайт с системой управления, два независимых DMA канала, последовательный и параллельный порты.

2.3.8. $n - \gamma$ разделение в нейтронной спектрометрии

В нейтронной спектрометрии для $n - \gamma$ разделения обычно применяется жидкий сцинтиллятор NE213 [33]. Истинные спектры нейтронов и γ -излучения могут быть разделены, если известна функция отклика $A(E, L)$ на моноэнергичный входной спектр. При облучении детектора потоком нейтронов или γ -квантов со спектром

$\Phi(E)$ измеряемый спектр излучения от сцинтиллятора

$$N(L)dL = \int_{E_{low}}^{E_{high}} A(E, L)\Phi(E)dE. \quad (54)$$

Это уравнение представляет собой интегральное уравнение Фредгольма первого рода, а задача восстановления исходного спектра является примером некорректно поставленной задачи. Методы решения таких уравнений широко описаны в литературе, и опуская подробности, можно сказать, что уравнение (54) сводится к решению матричного уравнения

$$N = A\Phi, \quad (55)$$

так что исходный спектр Φ получается при обращении матрицы A :

$$\Phi = A^{-1}N. \quad (56)$$

Элементы матрицы A известны с некоторой ошибкой, к тому же она является плохо определенной, поэтому так просто матрицу обратить нельзя. Решение уравнения (54) находится с помощью методов регуляризации, и в вычислительном плане это довольно трудоёмкая процедура.

Для восстановления исходного спектра в работе [33] была использована программа FERDOR, минимизирующая величину ϵ^2 :

$$\epsilon^2 = \sum_{i=1}^{N_{row}} w_i (N_i - \sum_{j=1}^{N_{col}} A_{ij}\Phi_j)^2 + X, \quad (57)$$

где X учитывает априорное знание о спектре — решение должно быть положительно.

Альтернативным подходом является решение интегрального уравнения с помощью ANN. Для этого применяется полностью связанная ANN без скрытых слоёв, с линейной передаточной функцией и без внутренних уровней. Такая простая сеть может выполнять только ограниченный класс операций

$$\Psi = wN, \quad (58)$$

где N и Ψ — входной и выходной векторы, а w — матрица весов ANN. Сравнение уравнений (56 и 58) показывает, что они эквивалентны, если $w = A^{-1}$.

Сеть обучается минимизировать величину

$$E_p = \sum_j (\Psi_{pj} - a_{pj})^2, \quad (59)$$

где Ψ_{pj} — вычисленный выход узла j для входа узла p , а a_{pj} есть требуемый выход. Общая ошибка при обучении определяется как

$$E = \sum_p E_p. \quad (60)$$

Применяемая простая архитектура сети предполагает слабую корреляцию между входными узлами. Поэтому чтобы повысить обобщающую способность сети и упростить процесс обучения, входные данные были подвергнуты предварительной обработке и продифференцированы по E , так что входной сигнал D для сети

$$D = -E\left(\frac{dN}{dE}\right), \quad (61)$$

где N есть функция отклика. Это позволило значительно уменьшить корреляцию между входами ANN.

Обучение сети занимало около 1 часа на PC386 (25 мГц), а собственно восстановление исходного спектра, включая сглаживание, 3 секунды. Программа FERDOR выполняла эту же работу за 60 секунд. Проведенное сравнение результатов показало, что они хорошо согласуются во всей области интервалов по энергии отклика детектора.

2.4. Реконструкция треков

Для реконструкции треков в ФВЭ используются нейросетевые алгоритмы, разработанные Денби (Denby) и независимо Петерсоном (Peterson) (см. работы [4], [47]). Нахождение треков основано на применении нейронных сетей, имеющих обратную связь (FBNN). Для решения этой задачи используются MFT метод и метод эластичных сетей (elastic tracking), носящий еще название подхода деформирующихся шаблонов (deformable templates approach).

Нейросетевой подход, основанный на MFT более сильный, чем требуется для решения данной задачи. Метод эластичных треков оптимальнее MFT. Он использует априорные данные о форме треков, включает в себя процедуру фитирования, дающую информацию об импульсе частицы, образующей трек, и является менее чувствительным к шуму. Эффективность ANN алгоритмов около 99%, что сравнимо с эффективностью обычных методов (99,7%). На большом числе треков (> 100) ANN метод работает быстрее, чем обычные методы.

Заряженные частицы, возникающие в процессе взаимодействия, регистрируются с помощью трековых детекторов. Характерная множественность заряженных частиц⁷ $\simeq 10^1 - 10^2$, а число плоскостей $\simeq 10 - 50$. Задача реконструкции трека состоит в том, чтобы провести линию в пространстве через N точек, отвечающую следующим трём условиям:

- 1) трек должен быть гладким,
- 2) каждая точка должна принадлежать только одному треку,
- 3) на треке не должно быть точек бифуркации.

Как видно из перечисленного, трек просто представляет собой “хорошую” в математическом смысле линию. Отметим также отсутствие требований, чтобы трек

⁷Например для SSC ожидалось бы $\simeq 4$ события каждые 100 – 250 нс и множественностью частиц 100 – 500.

начинался в точке взаимодействия или был частью окружности — эти требования не выполняются для событий с тяжелыми кварками и при работе трековой системы в неоднородном магнитном поле. Обычные способы восстановления 4-импульса, где значительное место и время занимает обычный комбинаторный перебор точек, становятся непригодными из-за экспоненциального роста времени с числом точек (кластеров).

Рассмотрим формулировку задачи реконструкции трека для решения с помощью нейросетевого подхода [19], [48], [8]. Определим *сегмент* как линию, соединяющую две соседние точки из набора N измеренных. Для набора сегментов определим энергетическую функцию

$$E = E^{cost} + E_1^{constrain} + E_2^{constrain}, \quad (62)$$

где первое слагаемое

$$E^{cost} = -\frac{1}{2} \sum_{ijkl} \delta_{jk} \frac{\cos^m \theta_{ijl}}{r_{ij} + r_{jl}} \quad (63)$$

учитывает гладкость трека и стремится к наименьшему значению при минимальной кривизне линии, а два других слагаемых

$$E_1^{constrain} = -\frac{\alpha}{2} \left[\sum_{jk} S_{ij} S_{kj} + \sum_{jl} S_{ij} S_{il} \right] \quad (64)$$

и

$$E_2^{constrain} = \frac{\beta}{2} \left[\sum_{ij} S_{ij} - N \right]^2 \quad (65)$$

соответственно отсутствие точек бифуркации и равенство числа точек на треке общему числу измеренных точек, α , β — множители Лагранжа. Если идентифицировать сегмент с бинарным нейроном S_{ij} , то уравнение описывает энергию такой системы нейронов. Таким образом, задача свелась к минимизации энергии системы, описываемой функционалом. Существуют различные способы минимизации, но практически все они страдают тем, что решение может отвечать локальному минимуму. В случае нейронной сети конфигурация с минимальной энергией достигается либо методом *симулируемого охлаждения* (отжига), либо с помощью так называемой Mean Field Theory, когда веса нейронов обновляются по правилу

$$V_i = \tanh\left(\sum_j \frac{T_{ij} V_j}{T}\right), \quad (66)$$

где $V_i = \langle S_i \rangle_T$. Проведенные исследования показали устойчивую и быструю работу ANN, особенно в случае событий с большой множественностью треков.

2.5. Вычисление массы

Как уже упоминалось, традиционно ANN используются для задач распознавания и классификации образов. Но двоичное представление числа тоже можно рассматривать как образ, и тогда интересно попытаться научить ANN арифметическим действиям над числами, опираясь на ее способность к обобщению при обучении.

При поиске новых частиц обычно встречается проблема вычисления инвариантной массы по продуктам распада. В случае распада промежуточного бозона $W \rightarrow q\bar{q} \rightarrow hadrons$ ситуация осложняется тем, что адроны могут происходить не только от q -струй, но и от тормозного излучения самих адронов и глюонов. Рассмотрим метод обучения ANN для восстановления массы события по его адронным продуктам распада.

Для моделирования рассматривается детектор, аналогичный установке UA2 на SPS CERN. Реакция

$$p\bar{p} \rightarrow W \rightarrow q\bar{q} \rightarrow hadrons \quad (67)$$

сгенерирована программой RUTHIA. Адроны регистрируются и восстанавливаются в калориметре в переменных η, ϕ и E_{\perp} . Для дальнейшего анализа были отобраны события с не менее чем двумя струями и $E_{\perp} > 10$ ГэВ в области $|\eta| < 1$. Для обучающей выборки масса W -бозона была равномерно распределена в интервале $[40, 160]$ ГэВ, всего $3,6 \cdot 10^3$ событий. Работа ANN проверялась на двух различных наборах, в одном из которых M_W тоже была равномерно распределена в интервале $[50, 150]$ ГэВ, а в другом массы W - и Z -бозонов положены равными табличным (80 и 91,2 ГэВ соответственно). По работе на первой выборке можно оценить способность сети к обобщению, а вторая давала информацию о том, как ANN будет работать на реальных данных.

Архитектура ANN, разумеется, зависит от того, в каких переменных будет представлено событие. Всего было испытано 4 способа, начиная практически от “сырых” данных с калориметра — $E_{\perp}(\eta, \phi)$, 480 входных нейронов. Лучшие результаты были достигнуты при наиболее “интеллигентных” переменных, когда используется априорное знание о том, какая именно физическая информация полезна при восстановлении массы события. Процедура состояла в следующем: после реконструкции струй с помощью алгоритма LUCCLUS каждая описывалась 3 переменными, чувствительными к источнику струи,

- 1) n_{90} — число ячеек калориметра, суммарная энергия в которых составляет $\simeq 90\%$ энергии струи;
- 2) n_c — число ячеек с $E_{\perp} > 1$ ГэВ;
- 3) корреляция энергии в соседних счетчиках, определяемая как

$$s_c = \sum_{i,j} \frac{|E_{\perp,i}^2 - E_{\perp,j}^2|}{(E_{\perp,i}^2 + E_{\perp,j}^2 + 1)}, \quad (68)$$

где i и j — соседние ячейки калориметра.

Эти же величины были вычислены и для всего калориметра. Также были определены массы двух самых энергичных струй M_{12} , трех M_{123} и четырех M_{1234} ($M_{1234} = 0$ для событий с $N_{jet} < 4$). Общее количество входных переменных — 24, и все они были отнормированы на интервал $[0,1]$. FFNN состояла из двух скрытых слоёв и одного выходного узла с линейным выходом, который интерпретировался как число, а не как булевская переменная. Использование такого набора переменных

позволило существенно снизить общее количество нейронов в сети ($\simeq 10^3$ против $\simeq 10^5$ при выборе “сырых” переменных), что упрощает процесс обучения и увеличивает способность к обобщению. Обучение сети происходило с помощью *BPE* процедуры, минимизировалась величина

$$E = \sum_p \left(\log(M_o^{(p)}) - \log(M_t^{(p)}) \right)^2 = \sum_p \left(\log\left(\frac{M_o^{(p)}}{M_t^{(p)}}\right) \right)^2, \quad (69)$$

которая в случае малых ошибок примерно соответствует ширине распределения M_o/M_t :

$$E \simeq \sum_p \left(1 - \frac{M_o^{(p)}}{M_t^{(p)}} \right)^2, \quad (70)$$

где M_o — величина массы на выходе сети, M_t — требуемое значение массы для события p . Идея обучения состоит в том, чтобы стартовав с низкой температуры сети (большое усиление), по мере обучения повышать её для всё более точного восстановления массы события. На тестовой выборке $\simeq 10^4$ W -событий с равномерным распределением событий по массе работа ANN оценивалась по величине

$$\sqrt{\langle (M_W^{net}/M_W^o)^2 \rangle - \langle M_W^{net}/M_W^o \rangle^2}, \quad (71)$$

равной 0,15 для ANN и 0,19 в случае восстановления массы с помощью традиционного алгоритма восстановления струи LUCCELL.

Таким образом, показана способность ANN реконструировать массу W -бозона с результатом, превосходящим обычные способы, и распределение по реконструированной массе W более узкое и симметричное. Это говорит о том, что алгоритм правильно учитывает поправки на излучение глюонов. Развитый здесь подход будет играть важную роль при выделении процессов $H^o \rightarrow W^-W^+$ при энергиях LHC и подавлении фона из систем $W + jets$, $tt \rightarrow W + jets$.

2.6. Реконструкция продольного фазового пространства пучка заряженных частиц по спектру синхротронного излучения

Для контроля качества пучка в ускорителях заряженных частиц важна информация о распределении в фазовом пространстве сгустка частиц. Прямые измерения плотности распределения частиц в пучке не могут дать необходимой точности и полноты. Новый способ [66] основан на обработке синхротронного излучения, которое возникает при движении заряженных частиц в магнитном поле. Такой подход позволяет получить более точную информацию о внутренней структуре пучка.

Спектр синхротронного излучения от сгустка частиц имеет две выделенные области: область излучения со спектром, характерным для одной частицы (это излучение сконцентрировано в области малых длин волн), и область с относительно большими длинами волн, возникающими из-за излучения сгустка частиц как целого, так называемое когерентное излучение. Именно форма спектра в этих

двух частотных областях и несет информацию о пространственном и импульсном распределениях частиц пучка.

Интенсивность когерентного излучения может быть выражена в терминах излучения от одиночной частицы

$$dI_{coh}/d\omega = (N + N(N - 1)F(\omega))dI_s/d\omega, \quad (72)$$

где

$$F(\omega) = \left| \int \exp(i\omega(n \cdot r)/c)n(r)dr \right|^2, \quad (73)$$

где $n(r)$ — плотность распределения электронов в пучке, а $F(\omega)$ — формфактор этого распределения. Из этих уравнений видно, что вклад когерентности в спектр излучения сильно зависит от формы пучка, особенно при длине волны излучения сравнимой с размером сгустка пучка. Обычно, длина сгустка частиц намного больше критической длины волны $\lambda_c = 2\pi c/\omega_c$, так что в спектре синхротронного излучения существуют две области: низкочастотная ($\omega < 1,9 \cdot 10^7$ Гц), которая зависит только от распределения в импульсном пространстве частиц, и высокочастотная ($\omega > 2 \cdot 10^9$ Гц), определяемая пространственной структурой пучка. Поэтому анализ частотного распределения излучения позволяет получить информацию о таких параметрах пучка как длина сгустка, сорт и количество частиц в нем, импульс и др. На этих свойствах спектра синхротронного излучения и основан способ диагностики пучка.

Нейросетевой подход диагностики пучка основан на предположении, что зарегистрированный спектр может быть отображен непосредственно в указанные выше переменные. Для тренировки ANN было использовано аналитическое выражение интенсивности излучения в зависимости от распределения частиц в импульсном пространстве:

$$d^2 I/d\omega d\Omega = \int_0^\infty dp f(p) d^2 I_s/d\omega d\Omega, \quad (74)$$

где $f(p)$ — произвольная функция продольного импульса p , а $d^2 I_s/d\omega d\Omega$ — спектр от излучения одиночной частицы при угле $\theta = 0$. Распределение $f(p)$ представляется выражением

$$f(p) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(p - p_o)^2}{2\sigma_p^2}\right) \times \sum_o^\infty c_n He_n\left(\frac{p - p_o}{\sigma_p}\right), \quad (75)$$

где He_n — полиномы Эрмита, p_o — средний импульс и σ_p дисперсия импульса пучка.

Интенсивность излучения от одиночной частицы, включая когерентные эффекты

$$\frac{d^2 I_{tot}}{d\omega d\Omega} = \left(1 + (N - 1)F(\omega)\right) \frac{d^2 I_s}{d\omega d\Omega}. \quad (76)$$

Для получения численных значений спектра излучения уравнение (73) было численно проинтегрировано. Параметры пучка для генерации спектра типичны для

накопительного кольца SPring-8. Реперные точки были выбраны в областях, отвечающих когерентной части спектра ($1,9 < \omega(\Gamma u) < 1,9 \cdot 10^7$) и около критической частоты ω_c , $1,9 \cdot 10^9 < \omega(\Gamma u) < 1,9 \cdot 10^{14}$).

При генерации спектра состояние сгустка частиц задавалось шестью переменными:

- 1) N_{type} , тип пучка,
- 2) σ_l , длина сгустка,
- 3) N , число частиц в сгустке,
- 4) c_1, c_2, c_3 - коэффициенты в разложении по функциям Эрмита.

ANN состояла из 4 слоёв — входной, два внутренних и один выходной. Входной слой состоял из 10 узлов, каждый соответствовал реперной точке спектра. Для определения структуры пучка были использованы две сети с разной интерпретацией выходных узлов и одинаковой структурой: 10–20–6–3. Для ANN, анализирующей низкочастотную часть спектра, выходные узлы интерпретировались как σ_l, N, N_{type} . Для высокочастотной ANN выходные узлы отвечали величинам c_1, c_2, c_3 . Для определения весов ANN было сгенерировано 3000 синхротронных спектров, при этом менялась каждая из шести величин, описывающая состояние сгустка частиц. При тренировке сети было замечено, что сеть, обучающаяся на низкочастотной части спектра достигала минимума гораздо быстрее, чем “высокочастотная” ($20 \cdot 10^3$ и $80 \cdot 10^3$ итераций соответственно).

Таким образом, реализованный способ нейросетевой диагностики хорошо вычислял параметры пучка частиц, т.е. решал обратную задачу восстановления исходного спектра с хорошей точностью, недостижимой (по крайней мере, сегодня) при использовании других методов.

2.7. Обработка данных времяпроекционной камеры

В рамках подготовки эксперимента ICARUS была создана большая жидкоаргоновая времяпроекционная камера (ВПК). Сигнал каждого канала оцифровывается 8-бит параллельным АЦП с частотой 20 МГц (всего 2048 отсчетов). Физическая информация содержится в следующих параметрах измеренной последовательности [13]

- t_p время прихода переднего фронта сигнала, определяет координату трека в соответствующей плоскости;
- h_p амплитуда импульса, пропорциональная собранному заряду;
- τ_r время нарастания сигнала (передний фронт), которое зависит от угла между треком и плоскостью сбора заряда ВПК, а также от диффузии электронного облака во время дрейфа.

Из полученного массива данных сначала выделяется область, содержащая сигнал выше некоторого порога. Затем, для каждой локальной области, выполняется

фит методом наименьших квадратов, использующий аналитическое выражение для описания формы импульса

$$f(t) = h_p \frac{\exp\left(\frac{t_p-t}{\tau_d}\right)}{1 + \exp\left(\frac{t_p-t}{\tau_r}\right)} + a + bt. \quad (77)$$

Как легко видеть, оно содержит 6 параметров: h_p — амплитуда импульса; t_p — время прихода импульса; τ_r — время нарастания; τ_d — время отклика усилителя и линейное изменение базовой линии $a + bt$.

Применение ANN основано на идее, что сеть может осуществить однозначное соответствие между двумя наборами векторов — измеренной последовательности сигналов и физическими параметрами трека с достаточной точностью и быстротой, без фитирования в явном виде. Входными параметрами для сети являются последовательность оцифрованных значений амплитуд достаточно большой длины (должен помещаться весь импульс), а также время отклика усилителя и параметры изменения уровня базовой линии. Последние три величины допускают прямое измерение. Итогом обработки входной последовательности должны быть три параметра импульса: h_p, t_p, τ_r . Моделирование сети начиналось с архитектуры 102 – 200 – 3. В дальнейшем, с целью ускорения обучения, количество нейронов в промежуточном слое было уменьшено до 50.

Для обучения сети было сгенерировано 10^3 импульсов с распределениями основных определяющих форму параметров, близких к экспериментальным. На выходе программы генерации создавался массив из 100 слов, имитирующий результат оцифровки импульса ВПК параллельным АЦП. Алгоритм обучения — традиционный *BP*. Для достижения ошибки обучения $\simeq 10^{-2}$ требовалось $\simeq 2 \cdot 10^4$ эпох. Это отнимало 150 часов процессорного времени на рабочей станции DEC 3000/800.

Работоспособность сети была проверена на 10^4 событиях, полученных методом Монте-Карло, и на $5 \cdot 10^4$ реальных событиях. Восстановление амплитуды сигнала происходило с одинаковым разрешением во всем интервале и составляло 0,73 отсчета АЦП. Время нарастания сигнала тоже восстанавливалось с хорошей точностью, намного превосходящей искажения, возникающие из-за диффузионного размытия электронного облака. Пространственное разрешение (непосредственно связанное с точностью определения времени прихода сигнала) обученной сети было проверено на реальных событиях с космическими мюонами. Полученное разрешение (200 ± 3) мкм немного уступает разрешению, полученному с использованием фита (142 ± 3) мкм. Обработка полного события ВПК с фитом требует на порядок больше процессорного времени (22 секунды), чем при использовании обученной ANN (< 2 секунд), причем для сети скорость решения практически не зависит от количества узлов во внутреннем слое.

Обработка данных с помощью программного симулятора нейронной сети, реализованного на скалярном процессоре, имеет следующие преимущества по сравнению с обычной процедурой фитирования:

- более высокая производительность,
- процесс обработки может касаться только отдельных параметров, представляющих интерес,
- возможна настройка сети на любую форму сигнала, используя реальные данные.

Конечно, использование одного, пусть даже очень быстрого, процессора заметно сказывается на времени обучения. Поэтому авторы работы [13] планируют использовать многопроцессорный комплекс АРЕ-100.

2.8. Фитирование профиля пучка

В работе [1] рассматривается применение ANN в электронике переднего края (front-end) для задачи диагностики пучка.

Обычно ANN используются для задач распознавания образов, но они могут также рассматриваться как измерительный прибор, т.е. аналоговый выход интерпретируется как число.

Система диагностики пучка использует пластины из оксида алюминия, покрытые стрипами (полосками) золота. Для определения x - и y -положения пучка необходимы два отдельных детектора. Аналоговый сигнал со стрипов усиливается и подается непосредственно на вход ANN, выполненной на СБИС 80170NX. ANN имеет архитектуру $16 - 32 - (8 + 8)$. Выходы нейронов последнего слоя каждой группы суммируются.

Для обучения были сгенерированы пики в форме гауссианы, имеющие FWHM между 3-й и 8-й проволоками, и положением пика между 3-й и 13-й проволоками. Как ширина, так и положение пика представляли собой реальные цифры. Аналоговый вход СБИС $0 - 3$ отображался в интервал $[-1, 1]$. Восемь выходов использовались для определения FWHM и положения пика, что улучшает разрешение соответственно. Темп обучения сначала равен 0,2, но в ходе обучения уменьшался в $\simeq 20$ раз. После СІІ обучения среднеквадратичная ошибка (rms) уменьшалась до 0,01. Ошибка в определении положения пика составляла 0,1, т.е. одну десятую от ширины стрипа. В случае данных с шумом ANN показывает хорошую устойчивость и нечувствительность к помехам, когда их уровень составляет 10 – 15%.

Заключение

Изложенный выше материал показывает, какие преимущества может дать применение ANN в ФВЭ. Эти преимущества обусловлены прежде всего способностью ANN к разделению классов событий с довольно сложной границей между ними, использованию ассоциативной памяти (способностью работать с “неполными” и “размытыми” (fuzzy) данными), терпимостью к шуму, быстродействием ANN, способностью адаптироваться к новым условиям. Различные типы нейронных сетей

обладают разными свойствами. Так FFNN хорошо подходят для задач классификации, а FBNN — для задач оптимизации, для создания моделей ассоциативной памяти.

В условиях существенного снижения объёма финансирования физики высоких энергий и требованием максимального “режима экономии” мы считаем перспективным использование ANN в следующих задачах, решаемых в ИФВЭ:

1. *Обнаружение развивающихся отклонений в режиме работы дорогостоящего оборудования (клизтронов, магнитов, систем питания). Для этих целей могут быть использованы аппаратно-реализованные аналоговые ANN.*
2. *Разработка легко модифицируемых и адаптируемых программ реконструкции событий. Используя такое свойство ANN, как обучаемость на примерах, и опираясь на моделирующую программу, можно, по-видимому, ускорить разработку программ реконструкции.*
3. *Создание триггерных систем. Здесь ANN помогает быстро обработать большое количество снимаемой с детекторов информации.*

Безусловно, во многих случаях успешное применение ANN зависит от наличия специального электронного оборудования. Доступность и относительная дешевизна аппаратно-реализованных нейронных сетей способствует росту их популярности среди специалистов, работающих в области ФВЭ.

Мы надеемся, что данный обзор поможет заинтересованному читателю ориентироваться в том поистине огромном потоке монографий, статей, программных продуктов, посвященных ANN, а специалисты в области ФВЭ смогут найти в нём ссылки на работы, в которых самые различные задачи решались с помощью ANN.

Области применения нейросетевого подхода постоянно расширяются, и в среднем количество публикаций по этой теме удваивается каждый год. Чтобы ориентироваться в этом потоке статей и журналов, необходимо уметь пользоваться преимуществом, которое дает использование электронных средств информации. Много интересного можно узнать, если регулярно читать группу *comp.ai.neural-nets*, используя глобальную информационно-коммуникационную систему *Internet*. “Способ существования и правила хорошего тона” при работе в сети изложены и доступной для широкой публики форме в книге [68]. Различная интересная информация — статьи, обзоры, и программное обеспечение типа public domain (freeware) — хранится на *WWW* — серверах всемирной паутины.

В заключение авторы считают своим приятным долгом поблагодарить В.В.Бойко, А.И.Галушкина и В.Н.Цоя за поддержку данной работы.

Настоящая работа выполнена в рамках проекта Миннауки по “Перспективным Информационным технологиям” — “Применение нейронных сетей в физике высоких энергий”.

Список литературы

- [1] Akkila T. et al. Nucl. Instr. and Meth., 1993, v. A327, p.573

- [2] Alkon D. L. Memory Storage and Neural Systems, Scientific American, 1989.
- [3] Ethem Alpaydm. CAL: Networks that grow when they learn and shrink when they forget, International Computer Science Institute, Berkely, USA, May 1991.
- [4] Amendolia S. R. Neural Networks for Trigger, CERN School of Computing, 1992.
- [5] Simon Haykin. Neural Networks. A Comprehensive Foundation. Macmillan College Publishing Company, New York, 1994, 696pp.
- [6] Anselmo F. et al. CERN/ECP 93–5.
- [7] Alspector J. et al. Adv. Neural Inf. Proc. System 3, eds. Lippman et al. San Mateo, CA, USA (Morgan Kaufmann), p. 1015.
- [8] Athanasu G., Pavlopoulus P. and Vlachos S. Nucl. Instr. and Meth., 1993, v. A324, p. 320
- [9] Bethke S. et al. Nucl. Phys., 1992, v. B370, p. 310.
- [10] Клименко С. В., Лебедев А. А. Автоматизация сложных экспериментов в физике высоких энергий. В сб.: II Всесоюзный семинар по автоматизации научных исследований, Новосибирск. 1982.
- [11] Битюков С.И. и др. Препринт ИФВЭ 95-77. Протвино, 1995.
- [12] Catani S. et al. Phys. Lett. 1991, v. B269, p. 432.
- [13] Cennini P. et al. Nucl. Instr. and Meth. 1995, v. A356, p. 507.
- [14] Danielssen M. et al. Nucl. Instr. and Meth. 1994, v. A350, p. 322.
- [15] De Felice P. et al. Phys. Lett. 1995, v. B354, p. 473.
- [16] Denby B. et al. Nucl. Instr. and Meth. 1993, v. A335, p. 537.
- [17] Denby B. et al. Nucl. Instr. and Meth. 1993, v. A335, p. 296.
- [18] Lindblad Th. et al. Nucl. Instr. and Meth. 1993, v. A327, p. 603.
- [19] Denby B. Comp. Phys. Comm. 1988, v. 49, p. 429.
- [20] Duda R., Hart P. E. Pattern Classification and Scene Analysis, Wiley: New York, 1973.
- [21] Durbin R. and Willshaw D. An Analog Approach to the Traveling Salesman Problem using an Elastic Net Method, Nature 326, 689, (1987).

- [22] Fogelman-Soulie F. Neural Networks for Pattern Recognition: Introduction and Comparison to Other Techniques. New Computing Techniques in Physics Research II, World Scientific, 1992.
- [23] Freeman James A. and Scapura David M. Neural Networks: algorithms, applications and programming techniques. (Addison-Wesley, 1991)
- [24] Hertz J., Krogh A. and Palmer R.G. Introduction to the Theory of Neural Computation, Addison-Wesley, Redwood City, Ca. (1991).
- [25] Hecht-Nielsen Robert. Neurocomputing. Addison-Wesley Publishing Company, 1991.
- [26] Hornblad P. et al. Nucl. Instr. and Meth. 1993, v. A336, p. 285.
- [27] Howell J.A. Neural Network in Ion Source Control Environment, New Computing Techniques in Physics Research, DU CNRS, Paris, 1990.
- [28] Hultqvist K. et al. Nucl. Instr. and Meth. 1995, v. A364, p. 193.
- [29] Innocente V., Wang Y.F. and Zhang Z.P. Nucl. Instr. and Meth. 1992, v. A323, p. 647.
- [30] Клайнкнехт К. Детекторы корпускулярных излучений. Москва, МИР, 1990.
- [31] Kim Y. K. et al. Phys. Rev. Lett., 1989, vol. 63, p. 1772.
- [32] L3 Collaboration. Nucl. Instr. and Meth. 1990, v. A289, p. 35.
- [33] Koohi-Fayefh R. et al. Nucl. Instr. and Meth. 1993, v. A329, p. 269.
- [34] Alekhin S. I. et al. LBL-90 — A Guide to Experimental Elementary Particle Physics Literature, Lawrence Berkeley Lab, UC.
- [35] Lonnblad L., Peterson C. and Rognvaldsson T, Phys. Lett. 1992, v. B278, p. 181.
- [36] Lonnblad L. et al. Phys. Rev. Lett. 1990, v. 65, p. 1321.
- [37] Lonnblad L. et al. Nucl. Phys. 1991, v. B349, p. 675.
- [38] Ляпидевский В. К. Методы детектирования излучений. — М.: Энергоатомиздат, 1987.
- [39] Marchesini G. and Webber B. R. Nucl. Phys. 1988, v. B310, p. 461.
- [40] Mazzanti P. and Odorico R. Z.Phys. 1993, v. C59, p. 273.
- [41] McCord Nelson M. and Illingworth W. T. A practical Guide to Neural Nets. (Addison-Wesley, 1991).

- [42] McCulloch W. C., Pitts W.A. Bull. Math. Biophys. 1943, v. 5, p. 115.
- [43] Mutoh M., Oouma T., Shibasaki Y., Abe L. and Nakahara K. Development of a Diagnostic Sistem for Klistron Modulators Using a Neural Network, Proceedings of the International Conferens on Accelerator and Large Experimental Physiks Control Systems, KEK, Tsukuba, Japan, Nov.1991.
- [44] Окунь.Л. Б. Физика элементарных частиц. — М.: Наука, 1984.
- [45] Olds J. L., Anderson M. L., McPhie D. L., Staten L. D. and Alkon D. L. Imaging memory-specific changes in the distribution of protein kinase C within the hippocampus, Science 245:866-869,1989.
- [46] Перкинс Д. Введение в физику высоких энергий. — М.: Энергоатомиздат, 1991.
- [47] Peterson C. Rögnavaldsson T. An Introduction to Artiftial Neural Networks, Cern School of Computing, 1991.
- [48] Peterson C. Nucl.Instr. and Meth. 1989, v. A279, p. 537.
- [49] Peterson C. and Anderson J. R. A Mean Field Theory Leaning Algorithm for Neuural Networks, Complex Systems 1, 995 (1987).
- [50] Peterson C. and Anderson J. A. Neural Networks and NP-complete Optimization Problems; A Performance Study on the Graph Bisection Problem, Complex Systems 2, 59 (1988).
- [51] Методы анализа данных в физическом эксперименте. Ред. М.Реглер. — М.: МИР, 1993
- [52] Rumelhart D. E. and McClelland J. L. Parallel Distributed Processing: Exploration in the Microstructure of Cognition (Vol.1), MIT Press (1986).
- [53] Ribarics P., Fent J., Cruber A., Kiesling C., Oberlack H. Second Level Neural Network Trigger in the H1 Experiment, New Computing Technigues in Physics Research II, World Scientific, 1992.
- [54] Scientific American 3, 241, Sept. 1979, (Имеется перевод: “Мозг”, — М.: Мир, 1984).
- [55] Wasserman F. Practical Neurocomputing, 1990 (имеется перевод: Уоссермен Ф. Нейрокомпьютерная техника. — М.: МИР, 1992).
- [56] Х.Шмидт. Измерительная электроника в ядерной физике. — М.: МИР, 1989.
- [57] Stimpfl-Abele G., Yepes P. Comp. Phys. Comm. 1993, v. 78, p. 1.
- [58] 80170NX Electrically Trainable Analog Neural Network Data Booklet, Intel Corp. 22550 Mission College Boulevard, Santa Clara, CA 95052-8125, USA.

- [59] iNNTS Neural Network Training System Users's Guide, Intel Corp. (1992).
- [60] MA16 Programmable VLSI Array Prozessor for Neuronal Networks and Matrix-based Signal Processing. User Description, November 1993.
MA16 Development Software. User Manual. November 1993.
- [61] CLNN32 Experimental Research Prototype data sheet, Bellcore, Morristown, NJ 07960, USA.
- [62] Masa P. et al. Proc. Fourth Intern. Conf. on Microelectronics for Neural Networks and Fuzzy System, Turin, 1994.
- [63] Proc. of the Intern. Conf. on Computing in High Enrgy Physics'92 September 21-25, 1992, Annecy, France.
- [64] Proc. of the Third Intern. Workshop on Software Engineering, Artificial Intelligence and Expert System for High Energy and Nuclear Physics, October 4-8, 1993, Oberammergau, Germany, 351.
- [65] Pavlopoulos P. et al. Nucl. Instr. and Meth. 1995, v. A359, p. 566.
- [66] Nucl. Instr. and Meth. 1995, v. A363.
- [67] Leimburger F. R. et al. Nucl. Instr. and Meth. 1995, v. A365, p. 198.
- [68] Клименко С., Уразметов В. *Internet — среда обитания информационного общества*. Изд. РЦФТИ при поддержке РФФИ. — Протвино, 1995.

Рукопись поступила 29 августа 1996.

С.В.Клименко и др.

Искусственные нейронные сети в физике высоких энергий.

Оригинал-макет подготовлен с помощью системы \LaTeX .

Ответственный за выпуск коллектив авторов.

Подписано к печати 6.09.96. Формат $60 \times 84/8$.

Офсетная печать. Печ.л. 6,1. Уч.-изд.л. 4,7. Тираж 240. Заказ 741.

Индекс 3649. ЛР №020498 17.04.97.

ГНЦ РФ Институт физики высоких энергий
142284, Протвино Московской обл.

