

ИНДЕКС 3649

Препринт ЕФИ -1390(1)93

ԵՐԵՎԱՆԻ ՖԻԶԻԿԱՅԻ ԻՆՍՏԻՏՈՒՏ  
ЕРЕВАНСКИЙ ФИЗИЧЕСКИЙ ИНСТИТУТ  
YEREVAN PHYSICS INSTITUTE

---

А.А.Чилингарян, Р.М.Шагоян

Математические модели нейронных сетей  
как системы распознавания образов  
(часть II).



ЕРЕВАНСКИЙ ФИЗИЧЕСКИЙ ИНСТИТУТ



Ереван 1993

Ա.Ա. Զիլինգարյան Ռ.Մ. Շահոյան

ՆԵՅՐՈՆԱՅԻՆ ՃԱՆՅԵՐԻ ՄԱԹԵՄԱՏԻԿԱԿԱՆ ՄՈԴԵԼՆԵՐԸ  
ՈՐՊԵՍ ՊԱՏԿԵՐՆԵՐԻ ՃԱՆԱՉՄԱՆ ՀԱՄԱԿԱՐԳԵՐ  
(II մաս)

Ներկայացված են նեյրոնային ցանցերի՝ որպես պատկերների ճանաչման համակարգերի հետազոտությունների վերջին արդյունքները: Ուսումնասիրվել են Հոփֆիլդյան ցանցերի ասոցիատիվ հիշողության հատկությունները: Առաջարկվել են ուսուցման նոր եղանակներ, ստացվել են պատկերների ճանաչման սխալների եւ ուսուցման ժամանակի գնահատականները: Շեթավոր ցանցերի համար առաջարկվել է նոր նպատակային ֆունկցիա:

Խեցզետնաձեւ միզամածության տվյալների օրինակի վրա կատարվել է չերենկովյան պատկերների բազմաչափ անալիզի եւ նեյրոնային դասակարգման եղանակների համեմատությունը:

Աշխատանքը մասամբ կատարվել է 92 712 թեմայով:

Երեւանի Ֆիզիկայի ինստիտուտ

Երեւան 1993

Математические модели нейронных сетей  
как системы распознавания образов  
(часть II).

А.А.Чилингарян, Р.М.Шагоян

В работе приводятся последние результаты по исследованию и применению математических моделей нейронных сетей к распознаванию образов. Рассмотрены свойства рекуррентных нейронных сетей Хопфилдовского типа как систем ассоциативной памяти. Предложены новые алгоритмы обучения сетей, получены теоретические оценки ожидаемых ошибок классификации и времени обучения. Для последовательных нейронных сетей введена новая целевая функция.

На примере анализа данных наблюдения Крабовидной туманности Уипловским телескопом проводится сравнение методов многомерного анализа и нейронной классификации черенковских изображений.

Работа частично проведена по теме 92 712

Ереванский физический институт

Ереван 1993

A.A. CHILINGARIAN

R.M. SHAHOIAN

THE MATHEMATICAL MODELS OF NEURAL NETWORKS  
AS A PATTERN RECOGNITION SYSTEMS

The neural networks as a pattern recognition systems were investigated. Properties of Hopfield type NN as systems of associative memory were considered. The new algorithms of NN learning were suggested and also the estimation of learning time and expected misclassification were obtained. For layered NN the new quality function was introduced.

On the example of analysis of the observation data of Crab nebula the methods multivariable analysis and NN-classification of the Cherenkov's images were compared.

This work was supported in part by the grant 92 712

Yerevan Physics Institute  
Yerevan 1993

## 1. Введение

Основной особенностью нейронной информатики, как новой технологии вычислений, является представление решения задачи при помощи некоторой функционирующей во времени динамической системы. Модели динамических систем, релаксация которых может привести к осмысленным результатам (распознавание образов, принятие решений в условиях неопределенности и др.) используют некоторые особенности биологических нейронных сетей (многослойная архитектура, пороговость нейронов, нелинейность сигмоидной функции, свойство нейропластичности и др.). Построение разнообразных моделей нейронных сетей и их обучение решению множества прикладных задач позволяет с одной стороны получить более эффективные алгоритмы, а с другой - лучше понять особенности функционирования природного вычислительного устройства - головного мозга [1.1]. Для физиков изучение динамических процессов в больших, сильно связанных неравновесных системах, какими являются нейронные сети, представляет интерес само по себе [1.2].

Практически все исследуемые и используемые модели нейронных сетей связаны в основном с выбором некоторых частных структур с известными свойствами. В настоящей работе рассматриваются наиболее популярные модели - рекуррентные сети типа Хопфилда и многослойные перцептроны. Подробное описание структуры и динамики моделей приведено в первой части препринта [1.3]. Вторая часть посвящена в основном вопросам обучения нейронных сетей - ключевому элементу и биологических и искусственных сетей, наиболее трудно поддающемуся пониманию и моделированию.

Обучение Хопфилдовских сетей, заключается в такой модификации матрицы нейронных связей, что некоторые выделенные состояния сети (прототипы) оказываются инвариантными относительно динамики сети и, кроме того, эти состояния "притягивают" до определенной степени искаженные прототипы. Размеры этих областей притяжения ограничены из-за присутствия ложных аттракторов, представляющих собой сильно перемешанные различные прототипы и инверсные

к ним состояния, а также, из-за перекрытия областей аттрактивности, соответствующих альтернативным прототипам. Предложенный Э.Гарднер алгоритм итеративного обучения, приводящий к асимметричным матрицам связей, позволяет регулировать "силу" притяжения к прототипам (величину локального поля), однако вопрос о характере изменения области аттрактивности при увеличении параметра стабильности до сих пор остается в значительной степени открытым.

Во второй главе предпринята попытка количественного исследования и визуализации функционирования алгоритма Э.Гарднер в зависимости от значения управляющего параметра и степени искаженности распознаваемых образов. Проведена теоретическая оценка предельных значений параметров, ожидаемых значений ошибок распознавания и времен обучения.

Третья глава посвящена сравнительному исследованию одного класса обучающих алгоритмов последовательных сетей (многослойных перцептронов). Среди рассматриваемых четырех модификаций метода случайного поиска с возвратом при неудачном шаге, наиболее быструю сходимость обеспечивает вновь предложенный алгоритм, оперирующий с параметрами каждого отдельного нейрона (одной из линий в сложной разделяющей поверхности). Там же в качестве простого теста для сравнения различных методов обучения предложено использовать одно-, двух-, и трехместные булевы функции.

В четвертой главе рассмотрено применение нейронной техники к решению задачи обработки данных реального физического эксперимента - поиску локальных источников  $\gamma$ -квантов очень высоких энергий черенковским телескопом Уиппловской обсерватории (США). Предложенный новый тип целевой функции позволил апостериорно оптимизировать форму наилучшей области выделения сигнала ( $\gamma$ -кластер) и увеличить достоверность сигнала от Крабовидной туманности по сравнению с ранее применяемыми методами.

## 2. Исследование возможности управления областью аттрактивности в рекуррентных сетях

За последние годы моделям рекуррентных нейронных сетей Хопфилдовского типа [2.1] было посвящено немало работ, в основном исследующих общие вопросы определения максимальной емкости  $\alpha_c$ , ее зависимости от коррелированности образов [2.2] и от алгоритма построения синаптической матрицы [2.3, 2.4] и т.д. в термодинамическом пределе, т.е. когда число спинов  $N$  бесконечно велико.

В то же время, определенный интерес представляет изучение свойств систем с конечным числом спинов и вдали от точки насыщения памяти ( $\alpha < \alpha_c$ ). Так, например, биологические системы действуют вдали от точки насыщения. Кроме того, в большинстве практических задач стоит вопрос о различении двух альтернативных конфигураций: например, регистрация сигнала на уровне шума [2.5], когда эффекты, связанные с насыщением памяти, не влияют на построение матрицы взаимодействия  $\{J\}$ . Также актуальны вопросы распознавания образов при наличии конечного алфавита (прототипов). Отметим, что для случая двух образов, чисто комбинаторными методами можно рассчитать, каков будет процент ошибок при распознавании, и, таким образом, оценить тот или другой алгоритм с практической точки зрения.

С другой стороны, остается множество открытых вопросов, связанных с определением размеров областей аттракции и их управлением. Например: сколько областей аттракции существует в сети после ее обучения? Как зависит вероятность ошибочной классификации от алгоритма обучения сети?

В настоящей работе проводится изучение влияния корреляции между образами, степени искаженности образов (уровень шума) и, так называемого, параметра стабильности Э.Гарднер и некоторых других эффектов на качество обучения и распознавания образов.

Основное внимание уделяется исследованию и "визуализации" области аттрактивности образов (множество начальных конфигураций, релаксирующих к

обученному образу) и "управления" величиной этой области в процессе обучения.

Рассматривается стандартная модель Хопфилда из  $N$  полностью связанных спинов ( $\sigma_i = \pm 1$ ) с релаксацией при нулевой температуре

$$\sigma_i(t+1) = \theta(h_i(t)), \quad (2.1)$$

где  $\theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases}$  - ступенчатая функция а  $h_i(t)$  - локальное поле, действующее на  $i$ -тый спин на  $t$ -том шаге.

$$h_i(t) = \sum_{j \neq i} J_{ij} \sigma_j(t), \quad (2.2)$$

Обучение сети состоит в поиске такой матрицы взаимодействия  $\{J\}$ , чтобы заранее выбранные образы  $\{\xi\}$  оказались стабильными точками системы, т.е.

$$\sigma_i(t+1) = \sigma_i(t)$$

Саму релаксацию системы, в принципе, можно проводить тремя путями:

- 1) обновлять все спины системы в соответствии с условием (1) одновременно (параллельная динамика);
- 2) совершать обход системы в каком либо определенном порядке последовательно;
- 3) выбирать обновляемые спины случайным образом.

В своих расчетах мы пользовались третьим методом, т.к. первый часто приводит к осцилляциям, а для второго нет никаких естественных причин предпочтения того или иного порядка обхода другому.

Для построения матрицы  $\{J\}$  было предложено несколько алгоритмов, которые условно можно разделить на два класса.

В первом из них обучение проводится одновременно для всех связей и всех образов, присвоением связям  $J_{ij}$  определенных значений. Типичным представителем этого класса является алгоритм Хемминга [2.6].

$$J_{ij} = \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu, \quad (2.3)$$

где  $\mu$  нумерует образы, а  $P$  их число.

Этот алгоритм хорошо распознает лишь ортогональные образы, в связи с чем, для коррелированных образов был предложен т.н. псевдо-инверсный алгоритм [2.3]. (Численное исследование в [1.3])

$$J_{ij} = \sum_{\nu, \mu=1}^P \frac{1}{C_{\mu\nu}} \xi_i^\mu \xi_j^\nu; \quad C_{\mu\nu} = \sum_{i=1}^N \xi_i^\mu \xi_i^\nu / N. \quad (2.4)$$

Второй класс-это последовательно обучающие алгоритмы, когда связи последовательно обновляются

$$J_{ij}^{(t+1)} = J_{ij}^{(t)} + \Delta J_{ij} \quad (2.5)$$

вплоть до выполнения условия Э.Гарднер

$$\gamma_i^\mu = \xi_i^\mu \eta_i^\mu / |J_i| > G; \quad i = 1 \dots N, \quad \mu = 1 \dots p \quad (2.6)$$

$$|J_i| = \left( \sum_{j \neq i} J_{ij}^2 \right)^{1/2}.$$

Отметим, что хотя во многих алгоритмах понятие параметра стабильности отсутствует, качество распознавания определяется распределением  $\gamma_i$  (кроме того, важную роль играет еще и степень симметрии матрицы связей [2.7]).

В настоящей работе мы используем вариант алгоритма Гарднер, в котором связи обновляются согласно

$$\Delta J_{ij} = \begin{cases} \xi_i^\mu \xi_j^\mu & \text{if } G > \gamma_i^\mu \\ 0 & \text{if } G < \gamma_i^\mu \end{cases}. \quad (2.7)$$

На каждом шаге обучения для всех образов поочередно связи обновляются в случайно выбранном порядке. Обучение проводится вплоть до выполнения (2.6) для всех спинов и образов. Скоррелированность образов будем измерять расстоянием Хемминга.

$$d_H(\xi^\mu, \xi^\nu) = \left( 1 - \sum_{i=1}^N \xi_i^\mu \xi_i^\nu / N \right) / 2 \quad (2.8)$$

которое меняется от 0 для одинаковых до 1 для взаимно-негативных образов.

Численные расчеты проводились с помощью букв латинского алфавита на решетках 10x10 ( $\sigma = 1$  - клетка занята,  $\sigma = -1$  - клетка пустая). После обучения

матрицы, системе поочередно предъявлялись искаженные буквы с заданным заранее шумом  $\chi$  (перевернуто  $N$  случайно выбранных спинов). Вероятность ошибки  $P_{err}(G, \chi)$  вычислялась как отношение ошибочных распознаваний к общему числу предъявлений искаженных образов (100 экспериментов для каждой комбинации параметров, шум накладывался от 10 до 40%).

Отметим, что ошибки бывают 2-х родов:

1. Релаксация конфигурации, полученной из одного образа наложением шума, к другому образу.

2. Релаксация к так называемым ложным образам. В свою очередь ложные образы можно разбить на 2 вида:

первый вид - конфигурации, полученные линейной комбинацией обученных образов (и их негативов, которые так же являются фиксированными точками динамики системы);

второй вид - конфигурации, не связанные с исходными образами линейным преобразованием, которые, однако, так же удовлетворяют условию  $\gamma_i > 0$ . (Например, если связи  $J_{ij}$  после обучения имеют преимущественно одинаковый знак, система может релаксировать к состоянию ферромагнетика или антиферромагнетика в зависимости от знака связей).

Рассмотрим процесс обучения 2-х образов  $\{\xi^1\}$  и  $\{\xi^2\}$  с нулевой магнетизацией ( $\langle \xi^{\mu} \rangle = 0$ ) и расстоянием Хемминга  $d < 1/2$ . Разобьем  $N$  спинов на две группы. В группе А, состоящей из  $N(1-d)$  спинов, собраны те из них, которые в  $\{\xi^1\}$  и  $\{\xi^2\}$  совпадают ( $d_A = 0$ ), в группе В ( $Nd$  спинов) собраны спины, имеющие противоположные знаки в различных образах (группа В образует некий подобраз в  $\{\xi^1\}$  и его негативов в  $\{\xi^2\}$ ). Знаком "+" помечены области с  $\xi = +1$ , а "-" - с  $\xi = -1$  в образе  $\{\xi^1\}$ . (Для удобства мы можем перенумеровать спины см.рис.2.1.

Обучение начнем с начальной матрицы  $\{J^{(0)}\}$  с нулевым средним ( $\sum_{i \neq j} J_{ij}^{(0)} = 0$ ) и единичной симметрией  $J_{ij}^{(0)} = J_{ji}^{(0)}$ .

При каждой итерации, согласно алгоритму (2.3) связи между спинами при невыполнении условия (2.6), в зависимости от их принадлежности к той или иной группе, изменяются на величины, указанные в таблице 2.1.

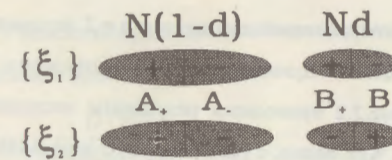


Рис.2.1

Таблица 2.1

$\Delta J_{ij}$	$A_+A_+$	$A_+A_-$	$A_+B_+$	$A_+B_-$	$A_-A_-$	$A_-B_+$	$A_-B_-$	$B_+B_+$	$B_+B_-$	$B_-B_-$
$\{\xi^1\}$	1	-1	1	-1	1	-1	1	1	-1	1
$\{\xi^2\}$	1	-1	-1	1	1	1	-1	1	-1	1

Отметим, что хотя сам метод обновления связей несимметричен, тем не менее, если изменяются  $J_{ij}$  и  $J_{ji}$ , то  $\Delta J_{ij} = \Delta J_{ji}$ . Это приводит к тому, что для больших  $G$ , когда на каждом шаге обучения меняется большинство связей, если абсолютные значения начальной матрицы  $|J| < 1$ , начальная симметрия теряется при первых же итерациях, а затем быстро возрастает до значений, близких к 1.

Из таблицы видно, что связи  $J_{ij}$ , связывающие спины из А и В при достаточно больших  $G$  (возможный выбор констант  $G$  обсуждается ниже), т.е. когда изменение затрагивает все связи, практически не изменяются. Это объясняет тот факт, что если в качестве начального приближения матрицы  $\{J\}$  взять большие абсолютные значения ( $|J| < 1$ ), то вероятность ошибок резко возрастает. Это связано с тем, что "необученные" связи типа  $J_{AB}$  сохраняют информацию о ложных образах (2-го типа), присущих  $\{J\}$ , которые и дают большую часть ошибок. В случае большего количества образов ( $P > 2$ ) связи между группами хоть и развиваются (в общем случае  $P$  образов систему можно разбить на  $2^{P-1}$  групп, хотя разбивка, в отличие от случая  $P = 2$ , неоднозначна), но хуже чем для спинов в одной группе в том смысле, что  $\Delta J_{AB}$ , определяемые различными образами, отличаются друг от друга по знаку.

С этой точки зрения, для коррелированных образов больше подходит псев-

доинверсный алгоритм, который для случая  $p = 2$  автоматически обеспечивает  $J_{AB} = 0$ . Конечно, для  $p > 2$  это не так, и разница между двумя алгоритмами уменьшается. На рис.2.2 приводятся результаты численных экспериментов с этими алгоритмами. Как видно, с увеличением  $G$  результаты алгоритма Гарднер улучшаются и сближаются с результатами псевдоинверсного алгоритма.

Очевидно, что спины группы В придется обучать дольше чем группы А (при  $d > 1/2$  -наоборот). Для спинов этой группы (для определенности возьмем спин из группы +) после  $t$ -той итерации получим

$$\xi_i h_i = \sum_{A+} J_{B+A+}^{(0)} - \sum_{A-} J_{B+A-}^{(0)} + \sum_{B+} (J_{B+}^{(0)} + 2t) - \sum_{B-} (J_{B+}^{(0)} - 2t) = h^{(0)} + 2tdN \quad (2.9)$$

$$|J_{B+}| \sim \left\{ \sum J^{(0)2} + (2t)^2 Nd + 4N(J_{B+B+}^{(0)} - J_{B+B-}^{(0)}) \right\}^{1/2}$$

где  $J_{B+}$  обозначает  $J_{B+A+}$ ,  $J_{B+A-}$  и  $J_{B+B-}$ .

Учитывая нулевую среднюю  $\{J^{(0)}\}$ , мы можем пренебречь линейным членом под корнем, и в результате получим

$$\gamma_{B+}^{(t)} = \frac{\xi_{B+} h_{B+}}{|J_{B+}|} \sim \frac{h^{(0)} + 2dNt}{\left( |J_{B+}^{(0)}| + 4t^2 Nd \right)^{1/2}} \quad (2.10)$$

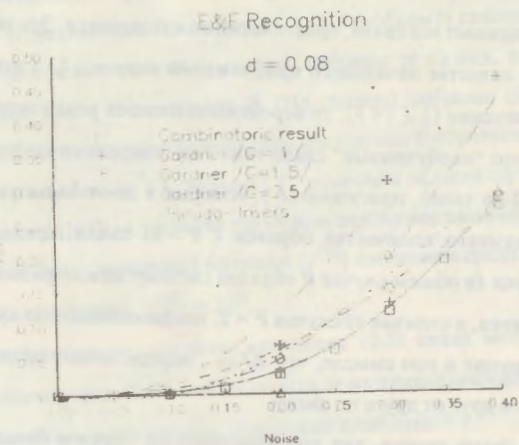


Рис. 2.2.

Из условия  $\gamma > G$  можно получить минимальное время обучения (число итераций)  $T$  и максимальное значение  $G_c$ , при котором алгоритм еще сходится.

$$T_{\min} \sim \frac{G |J^{(0)}|}{2(Nd(Nd - G^2))^{1/2}} \quad (2.11)$$

$$G_c = \sqrt{Nd} \quad (2.12)$$

При  $d > 1/2$  в обеих формулах  $d$  надо поменять на  $1-d$ , т.к. обучение некоторого образа эквивалентно обучению его негатива. Отметим, что реально обучение происходит дольше, чем (2.11), т.к. в процессе обучения условие  $\gamma_i^t > G$  для конкретного спина может то выполняться, то нарушаться (происходят осцилляции) и, таким образом, обновление  $J$  происходит не на каждой итерации. Тем не менее, (2.11) правильно объясняет тот факт, что при  $G \rightarrow G_c$  время обучения резко возрастает.

$$T(\epsilon) \sim \epsilon^2 / \sqrt{1 - \epsilon^2} \quad 0 \leq (\epsilon = \frac{G}{G_c}) < 1 \quad (2.13)$$

На рис.2.3 показана зависимость времени обучения  $T$  от  $G$ . На рис.2.4 показано

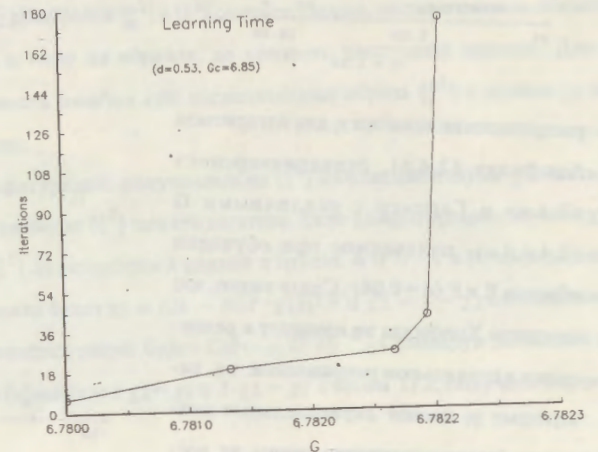
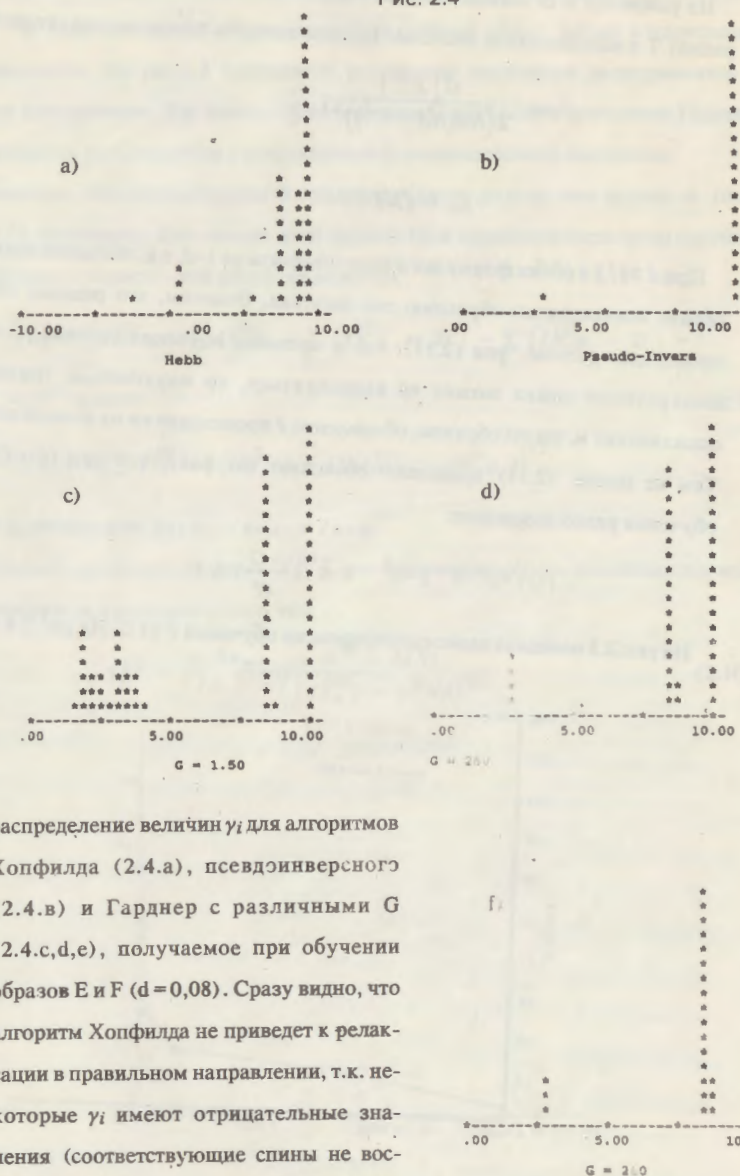


Рис. 2.3.

Рис. 2.4



распределение величин  $y_i$  для алгоритмов Хопфилда (2.4.а), псевдоинверсного (2.4.в) и Гарднер с различными  $G$  (2.4.с,д,е), получаемое при обучении образов Е и F ( $d=0,08$ ). Сразу видно, что алгоритм Хопфилда не приведет к релаксации в правильном направлении, т.к. некоторые  $y_i$  имеют отрицательные значения (соответствующие спины не восстановятся). Зато, псевдоинверсный ал-

горитм сразу создает 2 группы со значениями  $\sqrt{Nd}$  и  $\sqrt{N(1-d)}$ . Алгоритм же Гарднер, с увеличением  $G$  стремится собрать значения  $y$  к значениям, даваемым псевдоинверсным.

Заметим, что условие (2.12) сохраняется и при  $p > 2$ , просто в этом случае в качестве  $d$  надо брать расстояние между ближайшими образами. При  $G=G_c$  наступает ситуация, когда обновление связей при обучении спина из группы В образа  $\{\xi^1\}$  обязательно приводит к нарушению условия (2.6) для уже обученного спина из  $\{\xi^2\}$ . Конечно, (2.12) не единственное ограничение на  $G$ . Константа Гарднер ограничивается также и количеством образов  $P$ . Например для некоррелированных образов  $G_c$  и  $\alpha = P/N$  при больших  $N$  связаны соотношением [2.2]

$$\int_{-G_c}^{\infty} Dz (G_c + z)^2 = 1/\alpha; \quad Dz = \frac{\exp[-z^2/2]}{\sqrt{2\pi}}$$

Для коррелированных образов это выражение принимает гораздо более сложный вид.

В пределе  $G \rightarrow G_c$  при предъявлении системе конфигурации с расстоянием (шумом)  $\chi_1$  и  $\chi_2$  до образов  $\{\xi^1\}$  и  $\{\xi^2\}$  соответственно, система должна обязательно релаксировать к тому из образов, до которого расстояние меньше. Для этого случая вероятность ошибки при распознавании образа  $\{\xi^1\}$  с шумом  $\chi_1$  можно вычислить точно.

Из  $C_N^{N\chi_1}$  конфигураций, полученных из  $\{\xi^1\}$  наложением шума  $\chi_1$ , часть имеет меньшее расстояние до  $\{\xi^2\}$  или его негатива. Если конфигурация с шумом  $\chi_1$  была получена из  $\{\xi^1\}$  переворотом  $k$  спинов в группе А и  $N - k$  в В, то расстояние до  $\{\xi^2\}$  и его негатива будет  $\chi_2 = (2k - N(d - \chi_1))/N$  и  $\bar{\chi}_2 = 1 - \chi_2$  соответственно, и число таких конфигураций будет  $C_N^k (1-d)^{N-k}$ . Суммируя по всем  $k$  с  $\chi_2 < \chi_1$  и  $1 - \chi_2 < \chi_1$  и беря случаи  $\chi_2 = \chi_1$  и  $1 - \chi_2 = \chi_1$  с весом  $1/2$ , получаем вероятность ошибки

$$P_{err} = \left( \frac{C_N^{N(1-d)/2} C_N^{N(1-d)/2} + C_N^{N(1-d)/2} C_N^{N(1-d)/2}}{C_N^{N/2}} \right) + \left( \frac{\sum_k^{x_2(k) < x_1} \sum_k^{1-x_2(k) < x_1} C_N^{k(1-d)} C_N^{N-k}}{C_N^{N/2}} \right) \quad (2.14)$$

Соответствующая кривая (2.14) для образов E и F приведена на рис.2.2. Как видно, псевдоинверсный и алгоритм Гарднер при  $G \rightarrow G_c$  ( $G_c=2,68$ ) дают результаты, близкие к ожидаемым.

Попробуем объяснить зависимость количества ошибок от параметра стабильности. Пространство конфигураций сети из  $N$  спинов представляет собой  $N$ -мерный куб, который на рис.2.5 символически представлен сферой. Каждая точка сферы - одна из вершин гиперкуба. Матрице связей  $J_{ij}$ , при которой система всегда релаксирует к ближайшему образу, соответствует распределение областей аттракции, показанное пунктирной линией: вся поверхность сферы (гиперкуба) разделена на области, границы которых проходят посередине ближайших образов. Естественно, сюда входят и ложные образы 1-го рода: негативы и линейные комбинации образов (отмечены незаполненными буквами)

Однако, разброс значений величин  $\gamma_i$  в каждой из групп, полученных при обучении алгоритмом Гарднер, приводит к изменению этих границ (показаны сплошными линиями), и таким образом к зависимости качества распознавания от конкретной топологии шума в том смысле что некоторые спины оказываются обученными "хуже" чем другие. Кроме того, сюда добавляются еще и ложные образы 2-го рода (их области аттракции показаны штриховкой). Чем однороднее распределены величины  $\gamma_i$  в каждой из 2 групп (типа А, В...), тем глаже границы области аттрактивности каждого образа, как следствие локаль-

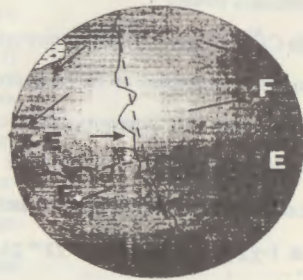


Рис. 2.5.

ной независимости, распознавания от конкретной топологии шума.

Для визуализации областей аттракции мы сгенерировали 1000 возможных конфигураций, отстоящих от образа F на расстоянии

$$x = \frac{\chi^{2F} - \chi^{2T} + d^2(F, T)}{2d} \quad y = \sqrt{\chi^{2F} - x^2}$$

(приведены только положительные  $y$ ), и его высота показывает количество конфигураций с такими шумами.

Граница "идеальной" области аттракции для F - это прямая  $x = 0,14$ , т.е. все конфигурации, лежащие левее этой прямой должны релаксировать к F.

Рис. 2.6.b,c,d показывает распределение ошибок при распознавании F с шумом 0.4 для матриц с различными значениями  $G$ .

Как видно из рис. 2.6.b, для малых  $G$  ( $G_c = 5.28$ ) большая часть ошибок лежит в области с  $x < 0.14$ . Это означает, что большая часть ошибок происходит из-за релаксации начальной конфигураций к негативным и ложным образам. Рис. 2.6.c,d показывает, что обучение с большими  $G$  позволяет избежать именно этих

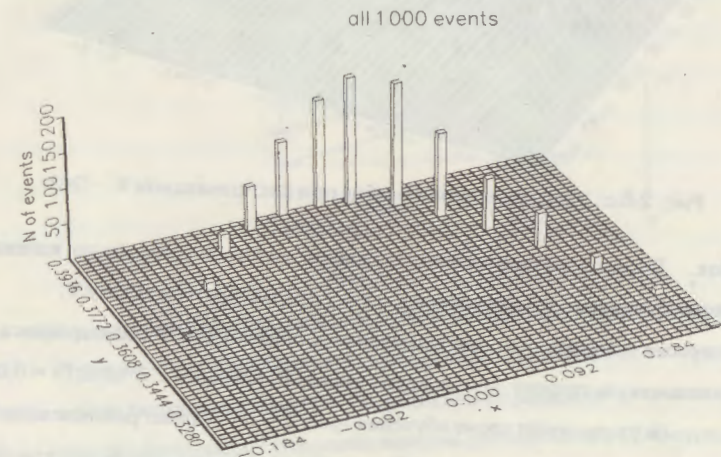


Рис. 2.6.a

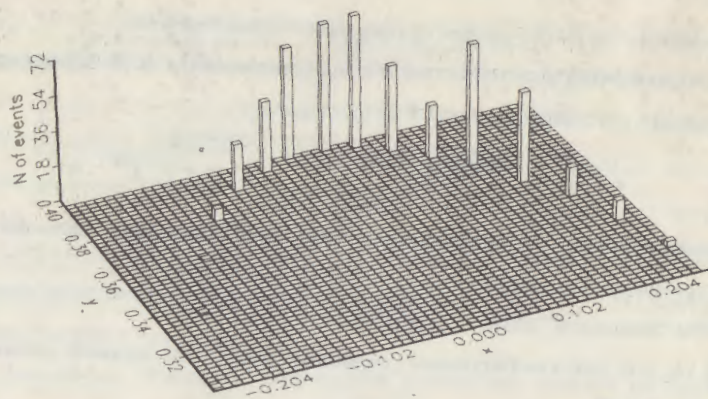


Рис. 2.6.b: Распределение ошибок при распознавании F. G=1.

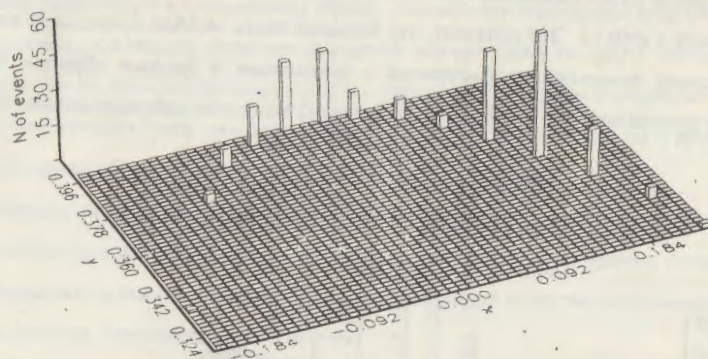


Рис. 2.6.c: Распределение ошибок при распознавании F. G=3.

ошибок. В конце концов остаются ошибки, которых невозможно избежать никаким обучением: конфигурации с  $\chi_F < \chi_T$  релаксируют к образу T.

Интересно отметить, что при добавлении к образам E и T, находящимся на значительном расстоянии друг от друга, образа F, близкого к E ( $d(E,F) = 0,08$ ), значительно увеличивает время обучения (рис. 2.7) и ухудшает распознавание не только образа E, но и "отдаленного" T (см. таблицу 2.2) - это увеличивает разброс

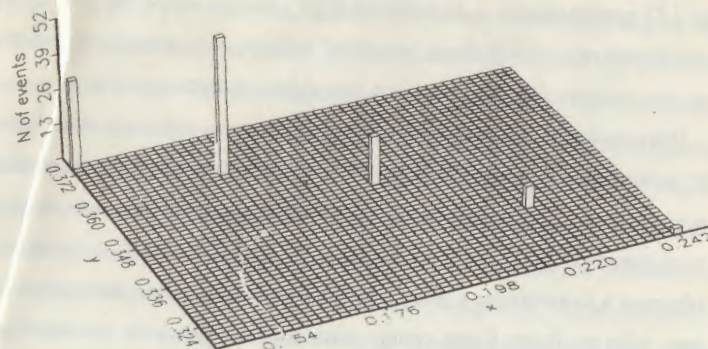


Рис. 2.6.d: Распределение ошибок при распознавании F. G=5.18.

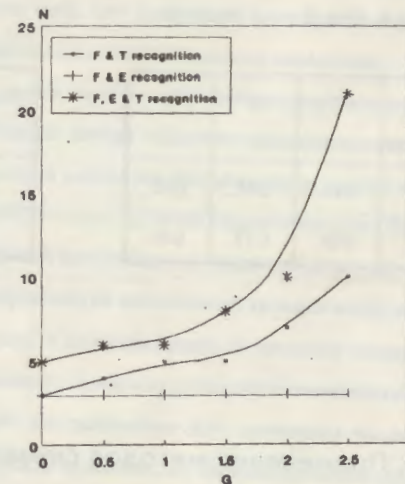


Рис. 2.6

в распределении и тем самым еще больше искажает границы области аттракции. Отметим, что такое поведение системы имеет аналогию в психологии: при выборе между двумя объектами (F и T) предъявление раздражителя, сходного с одним из них (E) приводит к ухудшению качества распознавания [2.8].

Еще одна интересная аналогия с реальной памятью заключается в том, что при

обучении, например 2-х образов с расстоянием  $d$ , увеличение  $\gamma_B^1$  на  $1/|J|$  прибавлением к  $J_{ij}$  величин  $\xi_i^1 \xi_j^1$ , уменьшит  $\gamma_B^2$  на  $(1 - 2d)/|J|$ , т.е. последний обученный образ "затирает" информацию о предыдущем. В результате, чем позднее был обучен образ, тем лучше он распознается.

Исходя из того, что для каждого образа существует свое максимальное значение  $G$ , можно предложить модификацию алгоритма Гарднер: брать не один общий для всех параметр стабильности  $G$  при обучении, а требовать для каждого образа выполнения  $\gamma_i^k > G^k$ , беря в качестве  $G = (N d_{\min}^k)^{1/2}$ , где  $d_{\min}^k$  - расстояние между образом и ближайшим к нему. Это приведет к тому, что распределение  $\gamma_i^k$  для всех образов будет более однородным в своих границах, повысив надежность обучения и независимость качества распознавания от конкретного распределения шума.

Таблица 2. а, b. G=1.5

Noise Perr	F	E	T
20	0.07	0.02	0.03
30	0.09	0.17	0.03

Noise Perr	E	T
20	0.0	0.0
30	0.02	0.02

### 3. Применение методов случайного поиска для обучения последовательных нейронных сетей.

Теоретическим обоснованием столь эффективных способностей последовательных нейронных сетей классификации образов является теорема о том, что трехслойная сеть с достаточно большим числом нейронов в скрытом слое способна производить многомерное проектирование в компактных подпространствах не-

прерывных или дискретных метрических пространств с любой заранее заданной точностью [3.1, 3.2] (естественно, после проведения соответствующего обучения). Задачу классификации можно рассматривать как частный случай такого проектирования, а именно, проектирование многомерных экспериментальных данных на конечный набор дискретных величин, представляющих собой индексы альтернативных классов. Естественно, что одна только теорема существования для построения классификатора недостаточна, необходимо иметь надежные и быстрые алгоритмы обучения. Под алгоритмами обучения мы будем понимать как методы выбора оптимальной конфигурации сети (например, числа нейронов в скрытом слое), так и - методы определение наилучших параметров сети (выбор вида целевой функции и нахождения глобального минимума). Следует отметить, что для многослойных персептронов целевая функция обычно является невыпуклой функцией параметров сети, что затрудняет нахождение эффективных алгоритмов, гарантирующих нахождение глобального минимума.

В настоящее время основным способом обучения последовательных сетей являются различные модификации метода обратного прослеживания ошибок (error backpropagation). В настоящей работе мы рассматриваем другой класс методов - алгоритмы случайного поиска в пространстве параметров сети. Нам представляется, что сравнение различных алгоритмов обучения представляет определенный интерес в свете все расширяющихся приложений техники нейроинформатики к решению производственных и научных задач. В качестве эталонной выбрана задача обучения распознаванию булевских функций (аналогично использованию стандартного гауссовского распределения для сравнения различных методов оценивания функции плотности вероятности). Булевы функции характеризуются наглядными таблицами истинности, возможностью безошибочного распознавания, легкостью усложнения задачи путем перехода к линейно неразделимым функциям или/и увеличением размерности функции (см. рис.3.1.a -OR и 3.1.b -XOR). Отметим, что хотя булевы функции могут распознаваться пороговыми нейронами, мы будем использовать сигмоидные нейроны, позволяющие создавать более общие и эффективные распознающие сети.

Рис. 3.1.a. OR

IN		OUT
0	0	0
0	1	1
1	0	1
1	1	1

Рис. 3.1.b. XOR

IN		OUT
0	0	1
0	1	0
1	0	0
1	1	1

Обучение сети методом случайного поиска можно реализовать двумя, принципиально различными методами

1. На каждом шаге обучения изменять все параметры одновременно, т.е. после прибавления ко всем параметрам сети приращения  $S \text{ Rand}(-1,1)$ , где  $S$  - максимальный шаг приращения и потом только оценивать целевую функцию с возвратом в исходную точку при неудачном шаге.

2. Выбирать случайным образом  $N_1 < N$  изменяемых параметров (в частности, менять по одному) и производить проверку.

Сразу отметим, что только первый метод дает гарантию сходимости алгоритма за достаточно большое время (естественно предполагается, что архитектура сети выбрана правильно, и максимальный шаг приращения выбран достаточно большим).

Для  $N$ -местных булевых функций, пространство входных конфигураций которых есть  $N$ -мерный единичный куб, при начальном значении всех параметров  $1/2$ , уже максимальный шаг  $S=1$  гарантирует сходимость, т.к. все решения являются  $N$  мерными плоскостями, пересекающими куб.

В то же время, изменение всех параметров одновременно несколько избыточно, т.к. хотя одна итерация в целом может оказаться неудачной, тем не менее, некоторые параметры в результате приращения могут попасть в области решения задачи, и это "достижение" будет потеряно в результате возврата в исходную точку.

Второй метод позволяет уменьшить эту избыточность, однако в этом случае есть опасность, что алгоритм никогда не завершится, во всяком случае в принципе, причем вероятность этого возрастает с уменьшением числа изменяемых параметров. Поясним это на 2-мерном примере (рис.3.2). На рисунке показан случай (довольно-таки нереальный), когда для попадания в область минимума необходимо изменение обоих параметров (пунктирная линия), тогда как любое изменение только одного из них ни к чему не приведет. Конечно, с увеличением размерности параметров сети вероятность такой конфигурации падает.

Численные исследования проводились на примере 2-х местной булевой функции XOR. Минимальной необходимой конфигурацией сети для этой задачи является трехслойный персептрон с двумя входными, двумя скрытыми в среднем слое и одним выходным нейронами, т.к. для решения задачи необходимо разделить пространство входных конфигураций минимум на три подпространства. На рис.3.3. показаны 2 семейства решений: первое дается сплошными линиями, второе - светлыми.

Проверялись три реализации алгоритма:

- 1) Изменение всех параметров одновременно (20 000 тыс. итераций)
- 2) Изменение только одного, случайно выбранного параметра (100 000 итераций)
- 3) Изменение параметров только одного, случайно выбранного нейрона (30 000 итераций)

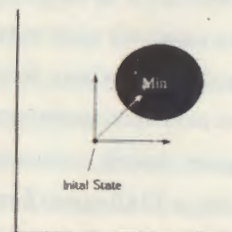


Рис. 3.2.

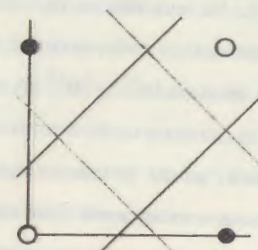


Рис. 3.3.

Наилучший результат показал 3-ий вариант: как видно из таблицы 3.1, он и сходиллся чаще (за разрешенное количество итераций), и допускал меньше неудачных шагов в пересчете на изменение одного параметра.

Таблица 3.1.

Метод	Процент удачных обучений	Число неудачных шагов на один удачный	То же самое в пересчете на 1 параметр
1	7	32	324
2	4	321	321
3	9	8	24

#### 4. Методы повышения достоверности наблюдения точечных источников $\gamma$ -излучения с помощью черенковских оптических телескопов

Задача достоверного определения точечных источников  $\gamma$ -квантов является самой важной в наблюдательной астрофизике очень высоких и сверхвысоких энергий. Черенковские телескопы с расположенным в фокусе кластером фотоприемников позволяют исследовать потоки  $\gamma$ -квантов в области промежуточных энергий  $10^{11} - 10^{13}$  эВ, недостижимой ни для спутниковых экспериментов из-за слишком слабых потоков, ни для крупных наземных скинтилляционных установок, ввиду затухания широких атмосферных ливней, инициированных частицами этих энергий. Однако черенковский свет от ШАЛ может быть собран оптическими рефлекторами и зарегистрирован с помощью кластера ФЭУ, расположенного в фокусе рефлектора.

Основным препятствием к применению черенковской техники является очень

большой фон от ШАЛ, инициированных протонами и ядрами космического излучения, который на 2-3 порядка превосходит ожидаемый полезный сигнал (от  $\gamma$ -квантов).

Традиционная методика проверки статистических гипотез заключается в принятии или отклонении так называемой нулевой гипотезы, которая заключается в отсутствии сигнала. То есть предполагается, что наблюдаемое превышение числа отсчетов при наведенном на источник телескопе, над числом отсчетов телескопа, смотрящего в "пустое" пространство,  $N_{on} - N_{off}$  обусловлено флуктуациями фона. Для доказательства факта существования источника необходимо отвергнуть эту гипотезу, и чем меньше будет вероятность справедливости нулевой гипотезы, тем выше будет достоверность вывода о наличии сигнала.

В условиях равенства времен наблюдения источника и контрольного наблюдения и при достаточно большом числе отсчетов в каждой серии наилучшим статистическим критерием является [4.1]:

$$\sigma = (N_{on} - N_{off}) / (N_{on} + N_{off})^{1/2} \quad (4.1)$$

где  $N_{on}$  - число отсчетов при наблюдении источника, а  $N_{off}$  - при наблюдении области звездного неба, где источники заведомо отсутствуют. Вероятность справедливости нулевой гипотезы равна интегралу от  $\sigma$  до  $\infty$  стандартного нормального распределения и она тем меньше чем больше  $\sigma$ .

Увеличения отношения сигнал/шум и значения  $\sigma$  можно достичь, произведя классификацию экспериментальных событий по какой-либо методике, с целью подавления фона от протонов и ядер космических лучей (сг), одновременно стараясь не очень сильно подавить полезный сигнал (изображения, инициированные  $\gamma$ -квантами).

Используя высокое качество изображений, полученных на 10-метровом рефлекторе Уипловской обсерватории [4.2], удалось перейти к рассмотрению каждого индивидуального экспериментального события (изображения). Параметры изображения вычисляются с помощью метода моментов распределения сигналов ФЭУ вокруг центра, восстановленного методом наименьших квадратов

эллипса потемнения.

Классифицируя экспериментальные изображения, можно значительно уменьшить число фоновых событий (одновременно будет потеряна и часть полезных событий). Новое значение критерия будет равно:

$$\bar{\sigma} = (\bar{N}_{on} - \bar{N}_{off}) / (\bar{N}_{on} + \bar{N}_{off})^{1/2} \quad (4.2)$$

где  $\bar{N}_{on}$ ,  $\bar{N}_{off}$  - число "выживших" после классификации событий.

В настоящее время в коллаборации Уипл проводятся работы по поиску наилучших дискриминирующих параметров изображения, по данным наблюдения Крабовидной туманности в 1988 - 1989 гг [4.3]. Эти данные состоят из 66 двадцативосьминутных ON/OFF сканов - всего - миллион событий, и достоверность факта наличия сигнала до распознавания изображений (по исходным данным с применением только аппаратурного триггера - 2 канала из 91 40 фотозлектронов) составило  $\sigma = 4.8$  (см. первую строку таблицы)

Проведение классификации изображений по параметру AZWIDTH, учитывающему как размер, так и ориентацию изображения, позволило в 4 раза улучшить достоверность ( $\sigma = 20$ ), и довести отношение сигнал/шум до 0.22 (вторая строка таблицы)

Таблица 4.1 Результаты применения различных методик подавления фона к данным наблюдения Крабовидной туманности Уипловским телескопом.

Метод анализа	$N_{on}$	$N_{off}$	Diff	$\sigma$	$\bar{\sigma}/\sigma$	$P_{\sigma \rightarrow \gamma}$	$P_{\gamma \rightarrow \gamma}$	$\frac{Diff}{N_{on}}$
Row Data	506255	501408	4847	4.8	-	-	-	0.01
AZWIDTH cut	14622	11389	3233	20.4	4.23	0.0227	0.67	0.22
Wedge cut	6017	3381	2636	27.2	5.67	0.0067	0.54	0.44
Supercut	4452	1766	2686	34.3	7.1	0.0035	0.56	0.60
Neural *	6278	2858	3420	35.8	7.5	0.0057	0.70	0.55

\*результаты получены по 1/4 экспериментальных данных, а затем пересчитаны

для полного экспериментального материала.

Дальнейшее улучшение достоверности наблюдения сигнала связано с переходом к многомерному корреляционному анализу черенковских изображений. Подход развитый в работах [4.4, 4.5], позволяет найти наилучшую комбинацию измеряемых параметров черенковского изображения и полностью использовать всю селективную (различительную) информацию.

Быстрота алгоритма многомерной классификации позволила провести априорную оптимизацию по всем экспериментальным событиям и выбрать наилучшие размеры многомерной области  $\gamma$ -изображений в трехмерном пространстве параметров AZWIDTH, WIDTH и LENGTH ( $\gamma$ -кластер) [4.6]. Решение о том, что анализируемое изображение принадлежит  $\gamma$ -кванту принималось при попадании 3 указанных параметров в заранее очерченную трехмерную клиновидную область (клин - WEDGE):

$$\begin{aligned} AZWIDTH &< 0.16 \text{ для зон } 2,3 \\ &< 0.15 \text{ для зон } 4,5 \\ LENGTH &< 0.3 \\ WIDTH/AZWIDTH &< 0.92 \end{aligned} \quad (4.3)$$

После этой демонстрации очевидных преимуществ многомерного анализа в коллаборации Уипл была проведена дальнейшая работа по оптимизации области  $\gamma$ -кластера. Была переведена в явную форму зависимость от прицельного параметра и ориентации изображения относительно центра поля зрения телескопа - был добавлен параметр DIST, а соотношение WIDTH/AZWIDTH заменено параметром ALFA. Таким образом, область  $\gamma$ -кластера представляет собой четырехмерный параллелепипед (метод назвали supercut [4.7]):

$$\begin{aligned} 0.510 < DIST &< 1.10 \quad . \quad 0.073 < WIDTH < 0.15 \\ 0.160 < LENGTH &< 0.30 \quad \quad 0.000 < ALFA < 0.26 \end{aligned} \quad (4.4)$$

Все большее увеличение критерия  $\sigma$  необходимо не только для получения большей достоверности существования сигнала от Крабовидной туманности, но прежде всего для поиска слабых, еще неизвестных источников и быстрого сканирования звездного неба.

Что и подтвердилось при открытии второго источника  $\gamma$ -излучения - эллиптической галактики Markarian 421, с применением параметров процедуры распознавания черенковских изображений, оптимизированных с помощью данных наблюдения Крабовидной туманности [4.8].

Однако оставался открытым вопрос: вся ли различительная информация используется и является ли выбранная форма  $\gamma$ -кластера (четырёхмерный параллелепипед) наилучшей?

Исследования многомерных распределений параметров черенковских изображений, попавших в область  $\gamma$ -кластера, проведенное М.Коули, позволило выявить значительные различия между ON и OFF сканами. С другой стороны, следует ожидать, что наилучшая область принятия решения имеет более сложную форму, чем прямоугольный параллелепипед.

Поэтому, дальнейшие попытки увеличить достоверность источника ведутся с использованием новой мощной дискриминационной техники - математических моделей нейронных сетей [1.3]. Нейронные сети представляют собой специальный класс статистических классификаторов, способных обучаться на предъявляемых примерах и, обеспечивающих сложные нелинейные области принятия решения.

Обучение сети производится путем обработки событий с известной принадлежностью и последовательной модификацией коэффициентов сети. После каждого цикла итерационного процесса, вычисляется функция качества:

$$Q = \sum_{k=1}^K \sum_{i=1}^{M_k} (OUT_i^k - OUT_{true}^k)^2 \quad (4.5)$$

где  $OUT_i^k$  - значение выходного нейрона сети для  $i$ -го события  $k$ -го класса,  $M_k$  - число событий в обучающей выборке  $k$ -го класса,  $K$  - число классов.  $OUT_{true}^k$  - "истинное" значение выхода для событий  $k$ -го

выбирать таким образом, чтобы максимально раздвинуть альтернативные классы.

Обучив сеть с помощью фоновых изображений (OFF скан) и моделированных  $\gamma$ -событий, удалось получить довольно высокие характеристики подавления фона [4.9], однако при обработке реальных ONN, OFF пар значение  $\sigma$  оказалось меньше, чем полученное с помощью многомерного анализа. Это обстоятельство можно объяснить неизбежными упрощениями и неточностями моделирования, так что результаты полученные в рамках выбранной статистической модели, значительно ухудшаются при переходе к реальным экспериментальным данным. Поэтому нами была разработана новая методика обучения нейронных сетей, использующая только экспериментальную информацию. Используя коэффициенты сети, полученные с применением функции качества (4.5) как первое приближение, дальнейшее обучение сети проводится с использованием ON, OFF пар и новой функции качества - (4.2). То есть апостериорно оптимизируется форма  $\gamma$ -кластера.

Для сокращения времени обучения сети, было произведена преселекция событий - для обучения отбирались события попавшие в четырехмерный параллелепипед со сторонами в 1.5 раз большими чем приведенные в (4.4). После 20000 итераций удалось получить значения параметров сети, обеспечившую улучшение достоверности наличия источника, по сравнению с методом *supercut* (см. последнюю строчку таблицы).

Представление о форме 4-мерного нелинейного  $\gamma$ -кластера, полученном с помощью нейронных сетей, дают гистограммы его проекций на 2-мерные плоскости.

В заключение авторы выражают благодарность М.Коули, в сотрудничестве с которым создавались и применялись новые методы анализа информации с черенковского телескопа Уиппловской обсерватории.

# BEST GAMMA CLUSTER SHAPE

rejected events

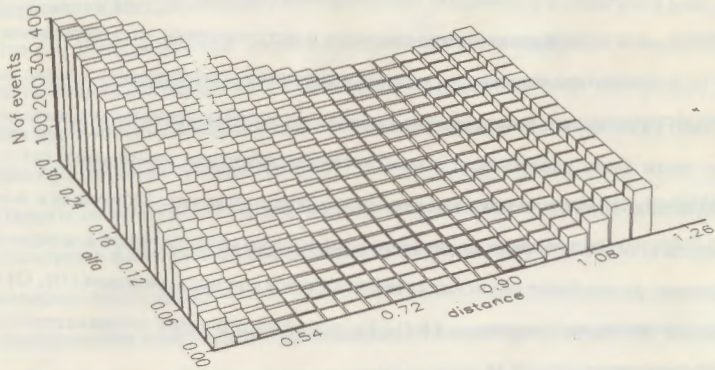


Рис. 4.1.a

# BEST GAMMA CLUSTER SHAPE

rejected events

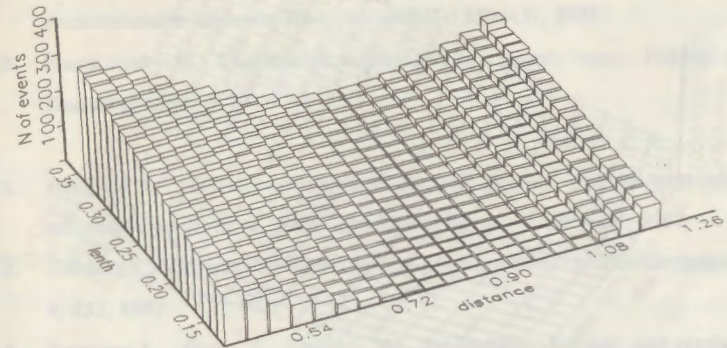


Рис. 4.1.c

# BEST GAMMA CLUSTER SHAPE

rejected events

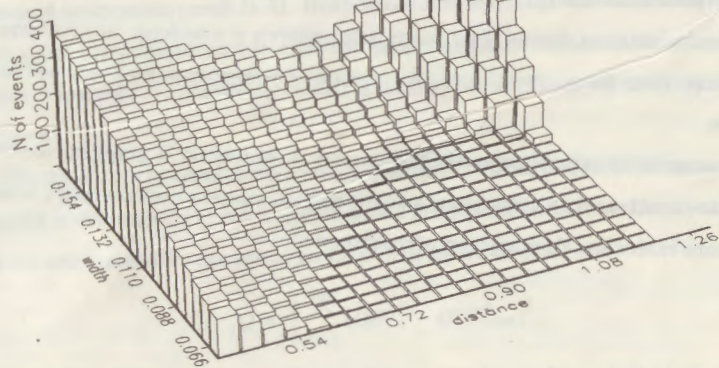


Рис. 4.1.b

# BEST GAMMA CLUSTER SHAPE

rejected events

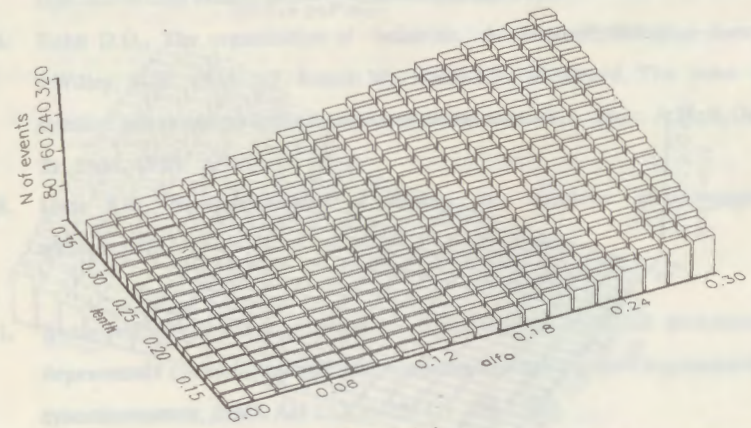


Рис. 4.1.d

## BEST GAMMA CLUSTER SHAPE

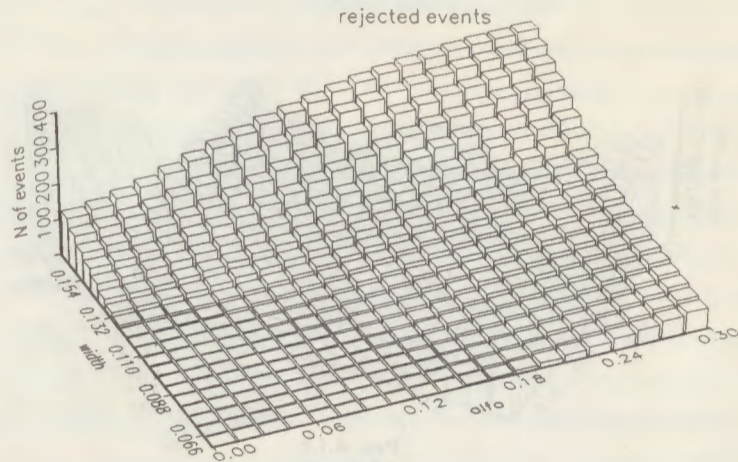


Рис. 4.1.e

## BEST GAMMA CLUSTER SHAPE

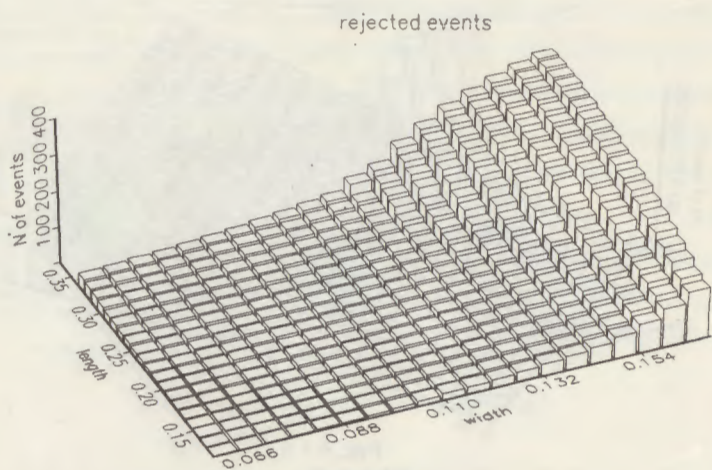


Рис. 4.1.f

## Литература

- 1.1. Галушкин А.И., Судариков В.А., Шабанов Е.В., Методика решения задач на нейрокомпьютере, Нейрокомпьютер, N 1, 1992 г., стр. 24-28.
- 1.2. Чилингарян А.А., Математические модели нейронных сетей как системы распознавания образов, Препринт ЕФИ-1350(45), 1991.
- 1.3. Sompolinsky H., The statistic mechanics of neural nets, Physics today, December 1988, p. 70.
- 2.1. Hopfield J.J., Neural networks and physical systems with emergent collective computational abilities, Proc. Nat. Acad. Scien. 79, 2554-58, 1982.
- 2.2. Gardner E., The space of interactions in neural network models, Europhys. Lett. 4, 451, 1987.
- 2.3. Persones L., Guon J., Dreyfus G., Information storage and retrieval in spin-glass neural networks, J. Physique Lett. 46, 359, 1985.
- 2.4. Abbot L.F., Learning in neural network memories, Preprint BRX-274, 1990.
- 2.5. Chilingaryan A.A., Neural classification technique for background rejection in high energy physics, Neurocomputing, 1993.
- 2.6. Hebb D.O., The organization of behavior. A neurophysiological theory, J. Wiley, N.Y., 1949.
- 2.7. Krauth W., Nadal J.P., M. Mezard, The roles of stability and symmetry in the dynamics of neural network, J. Phys. A: Math. Gen. 21, 2955, 1988.
- 2.8. Luria A.R., Neuropsychology of memory, M., Academy of pedagogical science, 1976.
- 3.1. Колмогоров А.Н., О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций одной переменной и суммированием, Докл. АН СССР, 1957, т.114, с.953.
- 3.2. Hecht-Nielsen R., Kolmogorov's mapping neural network existence theorem,

IEEE First Annual Int. Conf. on Neural Networks, v. 3, p.11, 1989.

- 4.1. Zhang S.N., Ramsden D., Statistical data analysis for  $\gamma$ -ray astronomy, Exp. Astronomy 1, p.145, 1990.
- 4.2. Cawley M.F., Fegan D.J., et al, A high resolution imaging detector for tev  $g$ -ray astronomy, Exp. Astronomy 1, p.173, 1990.
- 4.3. Vacanti G., Cawley M.F., et. al., Gamma - ray observations of the Crab nebula at Tev energies, Ap. J. 377, p.467, 1991.
- 4.4. Chilingarian A.A., Statistical decisions under nonparametric apriori information, Comp. Phys Comm. 54, p.381, 1989.
- 4.5. Aharonian F.A., Chilingarian A.A., et al, A multidimensional analysis of the Cherenkov images of air showers induced by very high energy  $g$ -rays and protons, Nucl. Ins.& Methods A302, p.522, 1991.
- 4.6. Chilingarian A.A., Cawley M.F., Application of multivariate analysis to atmospheric cherenkov imaging data from the crab nebula, Proc. 22 ICRC v.1, p.460, Dublin, 1991.
- 4.7. Punch M., Akerlof C.W., Cawley M.F., et. al., Supercuts: an improved method of selecting gamma-rays, Proc. 22 ICRC v.1, p.464, Dublin, 1991.
- 4.8. F.Halzen, Photons from hell, Nature, August, 1992.
- 4.9. Chilingarian A.A., Neural net classification of the  $\gamma$ - and  $p$ - images registered with atmocpheric cherenkov technique, random search learning in feed - forward networks, Proc. 22 ICRC v.1, p.540, Dublin, 1991.

The manuscript was received April 16, 1993

А.А.ЧИЛИНГАРЯН, Р.М. ШАГОЯН  
МАТЕМАТИЧЕСКИЕ МОДЕЛИ НЕЙРОННЫХ СЕТЕЙ  
КАК СИСТЕМЫ РАСПОЗНАВАНИЯ ОБРАЗОВ

Редактор А.С. Есян

Технический редактор А.С. Абрамян

---

Подписано в печать 28/IV-93

Формат 60x84x16

Офсетная печать. Уч.изд.л. 1,5

Тираж 100 экз.

Зак. тип. 125

Индекс 3649

---

Отпечатано в Ереванском физическом институте

Ереван-36, ул. Братьев Алиханян 2.