

2008-337

**НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК
КЫРГЫЗСКОЙ РЕСПУБЛИКИ**

ИНСТИТУТ ФИЗИКИ

Межведомственный диссертационный совет Д 01.07.342

На правах рукописи
УДК 681.327.68:778.38

САРЫБАЕВА АПЕЛ АКМАТБЕКОВНА

**ИССЛЕДОВАНИЕ ПРОЦЕССОВ ОБУЧЕНИЯ
НЕЙРОННЫХ СЕТЕЙ И ИХ РЕАЛИЗАЦИЯ**

Специальность: 01.04.05 – ОПТИКА

Автореферат

**диссертации на соискание ученой степени
кандидата физико-математических наук**

Бишкек 2008

Работа выполнена в лаборатории «Оптоэлектроника»
Кыргызско-Российского Славянского университета

Научный руководитель: доктор технических наук, профессор
А.А. Сагымбаев

Официальные оппоненты: доктор физико-математических наук,
профессор **З.Т. Азаматов**,
г. Ташкент, Узбекистан

кандидат технических наук, доцент
К.К. Талыпов, г. Бишкек, Кыргызстан

Ведущая организация: Алматинский инженерный центр по
лазерной технологии, Казахстан

Защита состоится «27» июня 2008г. в 14-00 часов на заседании Межведомственного диссертационного Совета Д 01.07.342 при Институте физики Национальной академии наук Кыргызской Республики, Биссыкульском государственном университете им. К.Тыныстанова и Ошском государственном университете, по адресу: 720071, г. Бишкек, проспект Чуй, 265-а.

С диссертацией можно ознакомиться в Центральной научной библиотеке НАН КР.

Автореферат разослан «22» мая 2008 г.

Ученый секретарь Межведомственного
диссертационного совета Д 01.07.342,
кандидат физико-математических наук *Мерен* - Л.К. Меренкова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В последние годы наблюдается существенный прогресс в развитии оптических систем обработки информации, достоинствами которых являются быстроедействие, параллельность обработки, широкий спектр осуществляемых в оптике операций над большими массивами данных. Среди таких систем особо важное значение имеют системы обработки информации на принципах голографической памяти. Целый ряд фундаментальных работ, среди которых важное значение имеют работы Д.Габора, Ю.Денисюка, П.Ван Хирдена, Х.Когельника, А.Микаэляна, А.Акаева, К.Жумалиева, А.Кутанова, С. Гуревича и многих других, позволили создать ряд эффективных систем голографической памяти. Однако, в связи с быстрым ростом требований к быстрдействию, эффективности и достоверности обработки больших массивов данных появляется необходимость дальнейшего совершенствования систем памяти. Большие возможности открывают методы обработки информации на принципах голографической памяти и нейронных сетей. Ассоциативность нейронных структур, так же как и голограмм, отличается возможностью поиска и считывания информации по любой её части. Нейронные сети и оптические системы взаимно дополняют друг друга, и их оптическая реализация является самым подходящим подходом для аппаратной разработки нейронной сети. Но их архитектура непосредственно связана с методом ее обучения. Даже различные этапы развития нейронных сетей определялись появлением новых архитектур сетей и специально разработанных для них методов обучения. Для обучения нейронной сети разработано много различных алгоритмов. Наилучшие показатели по скорости обучения среди всех алгоритмов обучения многослойных нейронных сетей дает алгоритм обратного распространения ошибки, который наиболее часто применяется при построении нейросетевых систем. Однако их основными недостатками являются локальность алгоритма обучения, высокая вычислительная сложность операций обучения и низкая устойчивость вычислительных процедур обучения. В настоящее время имеются различные модели оптических нейросетевых систем, и среди них двунаправленная ассоциативная память имеет много преимуществ: она совместима с аналоговыми схемами и оптическими системами; для нее быстро сходятся как процесс обучения, так и процесс восстановления информации; она имеет простую и интуитивно привлекательную форму функционирования. Но основным недостатком является ограниченная емкость памяти.

В связи с этим появляется необходимость дальнейшего развития методов обучения нейронных сетей, основанных на базовых принципах алгоритма обратного распространения ошибки и оптических нейросетевых систем на принципах двунаправленной ассоциативной памяти.

Целью диссертационной работы является: разработка новых высокоэффективных методов обучения нейронных сетей и создание оптических систем на принципах нейронных сетей.

Для достижения указанной цели необходимо решить следующие задачи:

1. Теоретически обосновать и выявить причины неустойчивости и вычислительной сложности математических методов обучения нейронных сетей;
2. Разработать новый метод обучения нейронных сетей для решения задачи с локальными минимумами;
3. Разработать новый метод для ускорения сходимости алгоритма обратного распространения ошибки;
4. Исследовать свойства двунаправленной ассоциативной памяти и на её принципах создать оптическую систему.

Научная новизна диссертационной работы заключается в том, что в ней впервые:

1. предложен алгоритм, позволяющий ускорить сходимость алгоритмов обучения нейронных сетей в случае попадания сети в локальные минимумы. Время обучения нейронных сетей с использованием алгоритма вывода сети из локального минимума сокращалось на 10-30%;
2. усовершенствован алгоритм обратного распространения ошибки для ускоренного обучения нейронных сетей за счет модифицированных функций активации: сигмоидной функции с различными параметрами, которые позволяют увеличить скорость обучения в 2 раза и синусоидальной функцией с дополнительными фазовыми параметрами, использование которой существенно расширяет возможности оптической реализации нейронной сети;
3. предложен новый подход нормировки данных с использованием функции активации, который позволяет повысить достоверность, надежность и эффективность нормировки путем приближения распределения данных к равномерному. Результаты экспериментов подтвердили эффективность предложенного способа нормировки, позволившего снизить ошибку обучения сети в 3-4 раза относительно данных, нормировка которых не проводилась;
4. разработана модифицированная модель двунаправленной ассоциативной памяти, у которой обучающая скорость в 1,5 раза быстрее и на её принципах разработана оптическая система. Результаты экспериментов показали чрезвычайно высокую помехоустойчивость системы, т.е. даже в условиях высокого внутреннего шума, система способна устойчиво работать и адекватно реагировать на соответствующие входные параметры.

Практическая значимость работы заключается в том, что использование предложенных алгоритмов обучения нейронных сетей позволяют сократить время обработки данных с помощью нейронных сетей и существенно расширяют возможности оптической реализации нейронных сетей; разработанная модель двунаправленной ассоциативной памяти, использующая усо-

вершенствованные алгоритмы, позволяет существенно увеличить помехоустойчивость оптической системы обработки информации на принципах нейронных сетей.

Полученные в диссертационной работе результаты использовались в научно-исследовательских и технических работах в учебном процессе КРСУ и в НАН Кыргызской Республики.

Основные защищаемые положения, выносимые на защиту:

1. Алгоритм, позволяющий ускорить сходимость алгоритмов обучения нейронных сетей в случае попадания сети в локальные минимумы;
2. Усовершенствованный алгоритм обратного распространения ошибки для ускоренного обучения нейронных сетей за счет модифицированных функций активации: сигмоидной функции с различными параметрами, которые позволяют увеличить скорость обучения и синусоидальной функцией с дополнительными фазовыми параметрами, использование которой существенно расширяет возможности оптической реализации нейронной сети;
3. Новый подход нормировки данных с использованием функции активации, который позволяет повысить достоверность, надежность и эффективность нормировки путем приближения распределения данных к равномерному;
4. Модифицированная модель двунаправленной ассоциативной памяти, на принципах которой разработана помехоустойчивая оптическая система.

Апробация работы. Основные результаты работы докладывались и обсуждались в ходе следующих конференций:

Международная конференция, посвященная 45-летию Кыргызского технического университета им. И. Раззакова (Бишкек, 1999), International Conference on Optoelectronics and Hybrid Optical/Digital Systems for Image Processing, (Львов, Украина, 1999), Международная конференция «Проблемы управления и информатики», (Бишкек, 2000), Международный семинар «Голография и оптическая обработка информации» (Бишкек, 1997, 2001), Международная конференция «Телекоммуникационные и информационные технологии. Состояние и проблемы развития», (Бишкек, 2002), Международная конференция «Развитие информационно-коммуникационных технологий в информационном обществе: состояние и перспективы», (Бишкек, 2005).

Личный вклад автора. Все основные научные результаты диссертационной работы получены автором лично. Экспериментальные исследования, относящиеся к настоящей работе, проведены совместно с Жэзибеквым А.А. и Сагымбаевым А.А.

Публикации. По теме диссертации соискателем лично и в соавторстве опубликовано 22 печатных работ.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, заключения, списка литературы 100 наименований, приложений изложенных на 160 страницах и содержит 45 рисунков и 8 таблиц.

СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность темы диссертации, сформулирована цель и задача исследования, определены научная новизна и основные положения, выносимые на защиту.

В главе I проведен обзор современного состояния теории нейронных сетей и техники их аппаратной реализации.

Анализ опубликованных работ показал, что настоящий прорыв в применении нейронных сетей для решения практических задач был сделан после того, как в 1986 году Румельхарт, Хинтон и Вильямс описали алгоритм обратного распространения ошибки. В настоящее время существует множество модификаций классического алгоритма обратного распространения ошибки. Однако существенной проблемой, возникающей при использовании данного метода является их локальность и сложность процедуры обучения нейронных сетей. В связи с этим появляется необходимость развития методов обучения нейронных сетей, основанных на базовых принципах алгоритма обратного распространения ошибки. А также, проанализировано состояние развития архитектуры нейронных сетей. Реализация нейронных сетей в виде оптических систем позволяет обеспечить важные преимущества при проведении вычислений. Рассматриваются различные модели оптических нейросетевых систем, и среди них двунаправленная ассоциативная память имеет много преимуществ: она совместима с аналоговыми схемами и оптическими системами; для нее быстро сходятся как процесс обучения, так и процесс восстановления информации; она имеет простую и интуитивно привлекательную форму функционирования. Но её основным недостатком является ограниченная емкость памяти. Представляет значительный интерес дальнейшее развитие модели двунаправленной ассоциативной памяти для реализации оптических нейросетевых систем.

Вторая глава посвящена исследованию свойств алгоритма обратного распространения ошибки для выявления причин его неустойчивости и вычислительной сложности и проведена оценка информационной емкости двунаправленной ассоциативной памяти.

Для исследования свойств алгоритма обратного распространения ошибки для обучения многослойных нейронных сетей исследуем поверхность ошибок (см. рис. 1).

Процесс обучения нейронной сети заключается в многократной подгонке весовых коэффициентов w_1, \dots, w_n к имеющимся выборочным данным с использованием различных методов нелинейной оптимизации. Алгоритм оптимизации начинает передвижение к минимуму E_{\min} , используя на каждом шаге вектор градиента (т.е. направления кратчайшего спуска). В общем случае

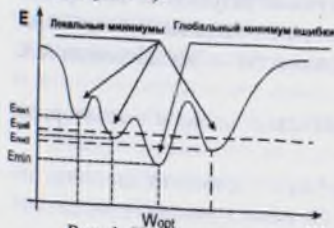


Рис. 1. Интерпретация поверхности ошибок

результат обучения нейронной сети может соответствовать субоптимальному решению, т.е. не глобальному оптимуму, а решению, которое нас устроит. В самом деле, если $E_{лок2} < E_{требуемое}$, то такое решение вполне приемлемо. Другое дело, когда веса попадают в область локального минимума (такого как $E_{лок1}$) и при этом $E_{лок2} > E_{требуемое}$, а величины шага обучения не хватает чтобы выйти оттуда. В ходе исследования были выявлены следующие факторы, влияющие на решение проблемы с локальными минимумами:

- проблема поиска глобального минимума ошибки;
- выбор шага обучения;
- выбор количества слоев и вида активационной функции.

Проведена оценка информационной емкости двунаправленной ассоциативной памяти, и результаты показывают, что:

- двунаправленная ассоциативная память имеет ограничения на максимальное количество ассоциаций, которые она может точно воспроизвести. Если этот лимит превышен, сеть может выработать неверный выходной сигнал, воспроизводя ассоциации, которым не обучена;
- двунаправленная ассоциативная память имеет ограниченную емкость памяти;
- сигмоидальная функция является оптимальной функцией активации благодаря ее способности усиливать низкоуровневые сигналы, в то же время, сжимая динамический диапазон нейронов.

В следующих главах представлены результаты решения указанных задач.

В главе 3 разработан новый алгоритм обучения нейронных сетей для ускорения сходимости в случае попадания сети в локальный минимум, усовершенствован алгоритм обратного распространения ошибки с учетом изменения функций активации, показано их влияние на сходимость алгоритма, а также предложен метод нормировки исходных данных.

Предложен новый алгоритм (блок-схема алгоритма представлена на рис.2) для ускорения работы алгоритмов при попадании нейронной сети в локальный минимум, целью которого является решение следующих двух задач: 1) обнаружить, что сеть попала в локальный минимум и скорость сходимости является низкой; 2) вывести сеть из локального минимума.

Вывод сети из локального минимума предложено осуществлять, изменяя значения весов и порогов сети на некоторые относительно небольшие величины таким образом, чтобы сеть сохранила весь предыдущий положительный опыт и в то же время могла изменить свое положение в многомерном пространстве весов и порогов. Заметим, что коррекция весов сети даже на достаточно малую величину может приводить к существенным изменениям, причем, не только положительным, но и к отрицательным, когда разработанный алгоритм ухудшает сходимость, достигнутую градиентным алгоритмом. Поэтому результаты изменений весов нейронных сетей на основе предложенного алгоритма должны приниматься только в том случае, когда они улучшают сходимость, в противном случае должны восстанавливаться значения весов,

полученные после применения корректирующего правила градиентного алгоритма.

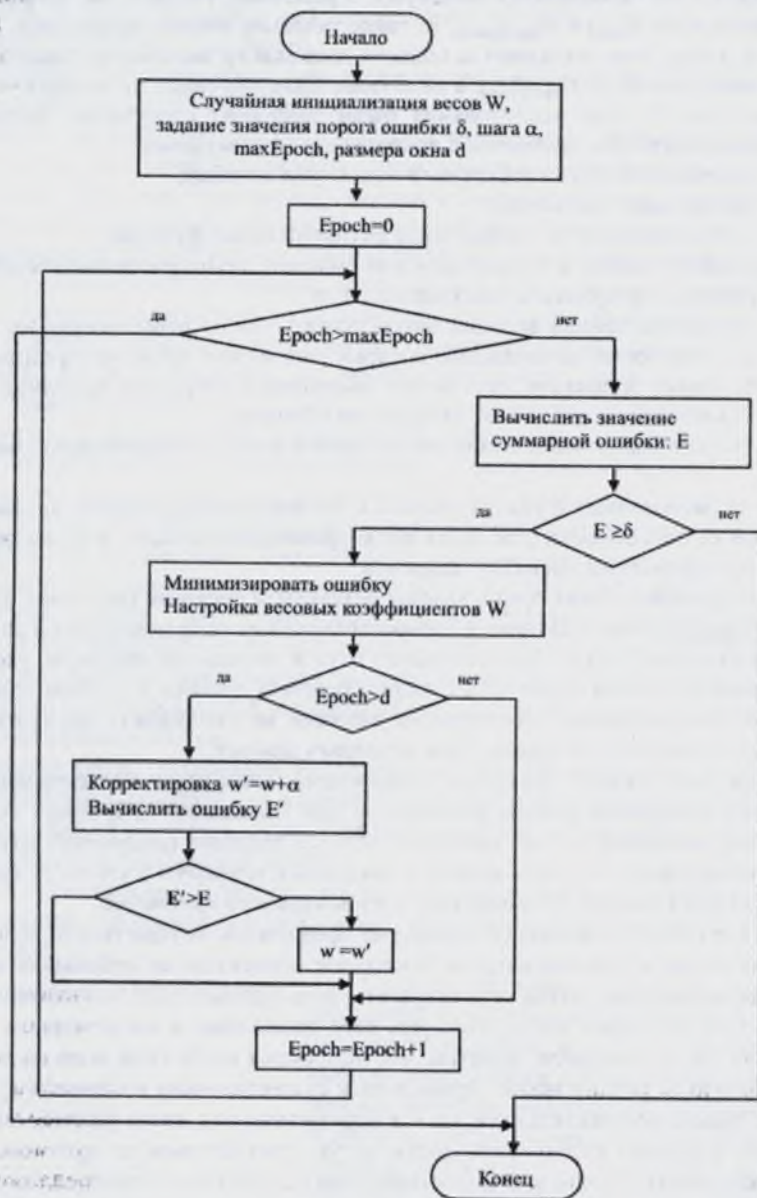


Рис. 2. Блок-схема алгоритма вывода сети из локального минимума

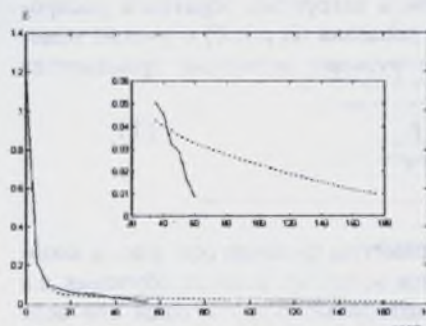


Рис. 3. Результаты тестирования нейронной сети со структурой 5-10-5

Для оценки эффективности предложенного и стандартного алгоритмов обучения создается многослойная нейронная сеть со структурой 5-10-5, предназначенная для классификации символов. Как показано на рис.3. предложенный алгоритм позволяет существенно сократить время обучения нейронной сети и увеличить скорость сходимости алгоритма. В проведенных экспериментах время обучения нейронных сетей с использованием предложенного алгоритма сократилось на 10-30%, что является очень

хорошим результатом.

Для того, чтобы выяснить, какие значения наиболее приемлемы для соответствующих параметров предложенного алгоритма проводились эксперименты, результаты которых показывают, что приемлемыми для большинства случаев могут быть следующие значения параметров: $\alpha = 1.03$.



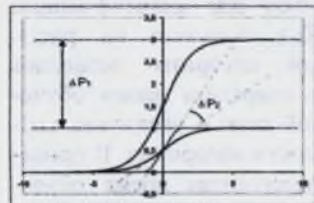
Рис. 4. Модель нейронной сети

Для корректировки шага изменения весовых коэффициентов рассматривается формула изменения весовых коэффициентов, для пары нейронов i, j (Рис. 4):

$$W_{ij}^{t+1} = W_{ij}^t + \eta \cdot E_j \cdot y_i^t,$$

где E_j – ошибка j -го нейрона, y_i – уровень активации i -го нейрона, а η – шаг изменения веса. Здесь η – величина постоянная, однако очевидно, что если он будет слишком мал, то обучение будет проходить очень медленно. Если же η велик – то, в момент достижения точки минимума (глобального или локального) функции ошибки, сеть не сможет в нее попасть и будет бесконечно долго колебаться вокруг нее, производя бесконечные пересчеты множества весов и ухудшая свои показатели. Поэтому, возникает необходимость управлять величиной η . Целесообразно в момент начала обучения нейронной сети установить некоторое его максимальное значение, обеспечив, таким образом, быстрое приближение к области, где находится W_{opt} и, затем, постепенно уменьшать его по приближению к самой точке оптимума: $\alpha_0 = 1$; $\alpha_{t+1} = \alpha_t - \delta\alpha$, где $\delta\alpha$ – коэффициент шага изменения весов сети. И опытным путем было установлено, что уменьшение α целесообразно производить, если $\Delta E < 0.1 \cdot 10^{-6}$, а $\delta\alpha = 0.001$. Использование такого условия позволило в течение всего процесса обучения сохранить достаточную скорость уменьшения ошибки.

Предложен метод ускорения сходимости алгоритма обратного распространения ошибки (блок-схема алгоритма показана на рис.6) с учетом изменения функций активации, т.е. в качестве функции активации применяется следующая функция:



$$y = \frac{p_1}{1 + e^{-p_2 x}}, \quad (1)$$

где, p_1, p_2 - параметры функции (см. рис.5), которые изменяются во время процесса обучения, а в стандартном алгоритме $p_1=p_2=1=const$ для всех элементов слоя.

Рис.5. График сигмоидальной функции

Для минимизации ошибки E и изменения параметров p_1 и p_2 сигмоидальной функции получены следующие выражения, соответственно:

$$\begin{aligned} \Delta W_{ij} &= -\eta_1 \frac{\partial E}{\partial W_{ij}} = -\eta_1 \frac{\partial E}{\partial y} \frac{\partial y}{\partial x} \frac{\partial x}{\partial W_{ij}}, \\ \Delta p_1 &= -\eta_2 \frac{\partial E}{\partial p_1} = -\eta_2 \frac{\partial E}{\partial y} \frac{\partial y}{\partial p_1}, \\ \Delta p_2 &= -\eta_3 \frac{\partial E}{\partial p_2} = -\eta_3 \frac{\partial E}{\partial y} \frac{\partial y}{\partial p_2}, \\ \Delta \varepsilon &= \frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial x} \end{aligned} \quad (2)$$

где, η_1, η_2, η_3 - наименьшие положительные константы,

$\Delta \varepsilon$ - изменение сигнальной ошибки для скрытых и выходных слоев нейронной сети.

Из результатов эксперимента видно, что изменение параметров p_1, p_2 сигма - функции существенно влияет на сходимость алгоритма.



Рис.7. Влияние длительности обучения на погрешность

Как показано на рис. 7, модифицированный алгоритм сходится к нужному решению при ошибке $E=0.0001$ через 8 с, т.е. позволяет увеличить скорость обучения в 2 раза. Отсюда можно заметить высокую эффективность и ошибкоустойчивость алгоритма обучения.

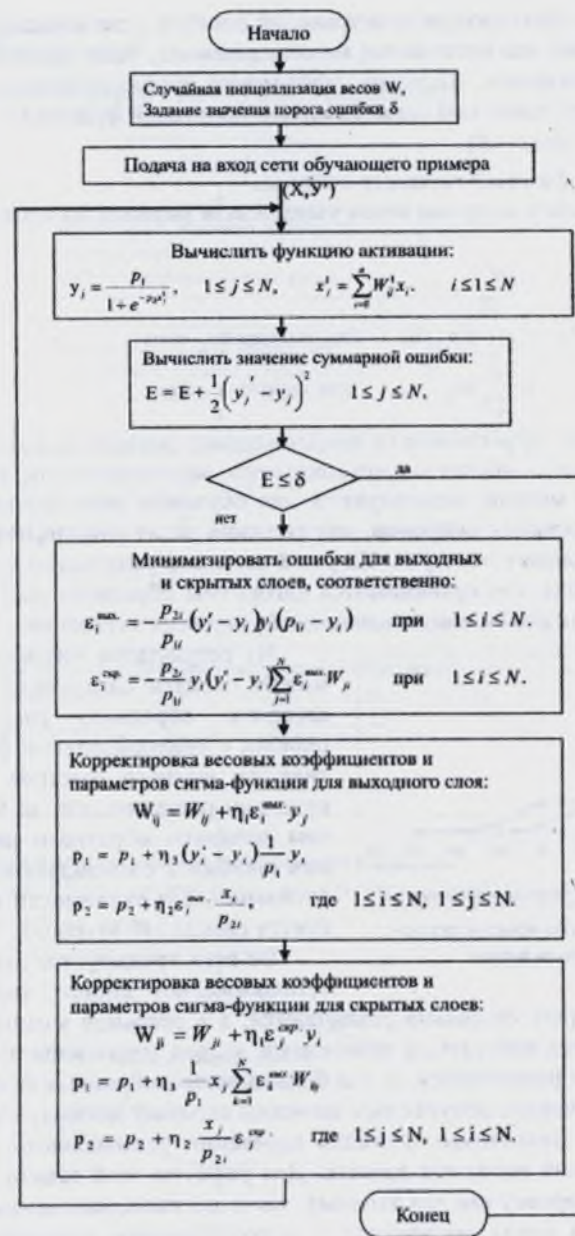


Рис. 6. Блок-схема алгоритма

Алгоритм обратного распространения ошибки с сигмоидальной функцией не применима при негативных межсоединениях, чаще применяемых в оптических реализациях. Поэтому предложим синусоидальную пороговую функцию вместо известной сигмоидальной пороговой функции:

$$f(x) = \sin(x - \theta), \quad (3)$$

где, θ - фазовый параметр нейрона.

Корректировка матрицы весов взаимосвязи выражается следующим образом:

$$\partial E / \partial W_{ij} = -\eta \sum_{i=1}^N \varepsilon_{ij} \cdot y_i, \quad (4)$$

где, $\varepsilon_{ij} = \begin{cases} 2[y_i - y'_i] \cdot f[x_j \cdot i], & \text{для выходного слоя} \\ f[x_j \cdot i] \cdot \sum_{q=1}^N W_{iq} \cdot \varepsilon_{iq}, & \text{для скрытого слоя} \end{cases}$

Для оценки эффективности предложенного алгоритма проведен эксперимент, где сеть обучается ассоциативно восстанавливать изображения. Предложенные методы используются для обучения двухслойной сети, состоящей из 4 скрытых нейронов, для решения задач классификации 8 четырехмерных бинарных векторов. На рис.8 показаны результаты компьютерного моделирования, где сравниваются алгоритмы обратного распространения ошибки с сигмоидной и синусоидальной функциями активации

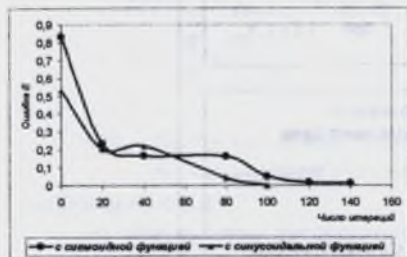


Рис. 8. Результаты компьютерного моделирования

Из результатов численного расчета можно сделать следующий вывод, что алгоритм обратного распространения ошибки с синусоидальной функцией активации намного быстрее сходится к нужному решению, т.е. за 98 итераций, чем алгоритм обратного распространения ошибки с сигмоидальной функцией активации, для сходимости которой требуется свыше 140 итераций.

Во всех предыдущих экспериментах использовались данные, диапазон представления которых не сильно различаются, а в реальной жизни нейронным сетям приходится работать со значениями входов, диапазоны представления которых сильно различаются. А для большинства нейронных сетей характерно наличие интервала допустимых значений входных данных, в пределах которого данные различимы. Функция активации устанавливает допустимые границы значений исходных данных. Для решения этой задачи необходимо провести нормировку как для входных, так и для выходных данных.

Предложен метод нормировки с использованием функции активации, выполняющейся по формулам:

$$s = f(x) = \frac{1}{1 + e^{-kx}} - p$$

где, k, p - используемая при обучении параметры.

(5)

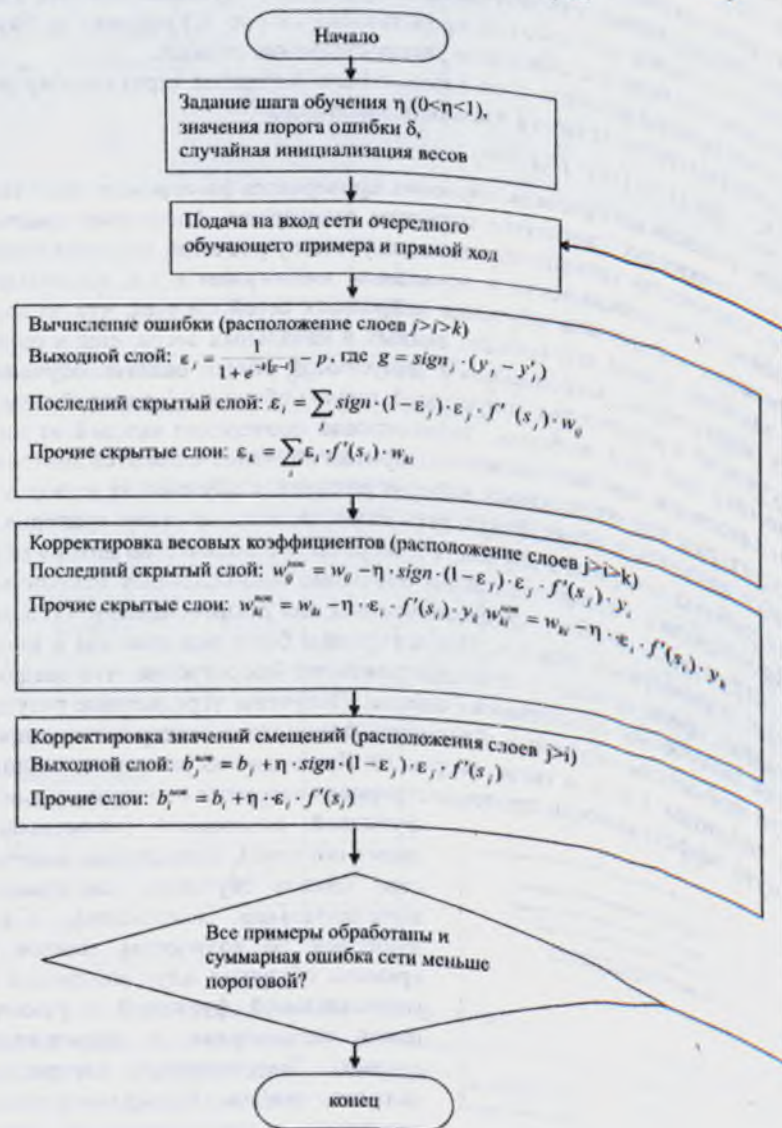


Рис. 9. Блок-схема алгоритма нормировки данных

Суть данного метода заключается в изменяемом виде нормирующей функции $f(x)$, что позволяет повысить эффективность нормировки путем лучшего приближения распределения данных к равномерному. Это позволяет увеличить информативность обучающих примеров и функции ошибки. Разработанный алгоритм (блок-схема представлена на рис. 9.) основан на базовых принципах алгоритма обратного распространения ошибки.

Ошибка предпоследнего слоя i может быть выражена через ошибку последующего (выходного) слоя j с помощью формулы:

$$\varepsilon_i = \sum_j \text{sign} \cdot (1 - \varepsilon_j) \cdot \varepsilon_j \cdot f'(s_j) \cdot w_{ij} \quad (6)$$

Эффективность алгоритмов обучения проверяется на определенных тестах, соответствующих принятым мировым стандартам. Алгоритмы сравниваются по количеству циклов обучения, количеству расчетов значения функции ошибки, чувствительности к локальным минимумам и т.п. Из анализа рассмотренных алгоритмов обучения нейронных сетей следует, что их сходимость зависит, кроме обучающих данных и начальных весов, еще и от задаваемых параметров: максимальной допустимой общей ошибки обучения (цель обучения) и количества допустимых циклов обучения (критерий скорости обучения) для всей выборки. Зафиксировав поочередно каждый из них, можем исследовать, как для данного алгоритма обучения меняются значения другого критерия при одинаковых наборах различных обучающих выборок и одинаковых начальных весах. Результаты были оценены по таким критериям, как время работы алгоритма обучения и затраченное количество циклов обучения. Проводились эксперименты по обучению многослойных нейронных сетей со структурой 5-10-5, предназначенный для аппроксимации 41 пары обучающих одномерных данных. Все алгоритмы были реализованы в инструментальной среде одного и того программного обеспечения, что создало основу для получения объективных оценок. Получены усредненные результаты по 20 процессам обучения. Фрагменты результатов экспериментов приведены в таблицах 1 и 2, а также на рис 10. Как видно из рисунка и таблиц, наибольшую эффективность продемонстрировал алгоритм с синусоидальной

функцией активации (наименьшее время обучения, наименьшее количество циклов обучения, наименьшая вычислительная сложность). Следующими по количеству циклов и времени обучения идут алгоритмы с сигмоидальной функцией с различными параметрами и нормировки данных. Предложенные алгоритмы являются наиболее быстрыми и менее сложными с вычислительной точки зрения, чем стандартный алгоритм.

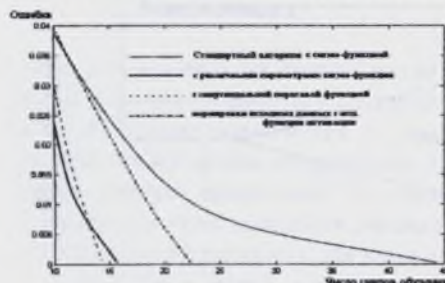


Рис. 10. Результаты сравнения эффективности алгоритмов

Таблица 1. Сравнительная характеристика алгоритмов обучения МНС при фиксированной максимальной точности

Алгоритм обратного распространения ошибки	Фиксированная максимальная точность (ошибка)					
	10^{-3}		10^{-3}		10^{-6}	
	Время обучения, с	Число циклов обучения	Время обучения, с	Число циклов обучения	Время обучения, с	Число циклов обучения
стандартный	2.03	4	9.05	32	10.11	45
с параметрами сигмоидной функции	1.91	3	7.64	18	6.32	15
с синусоидальной функцией	1.95	2	7.23	19	6.28	14
нормировки исходных данных	1.92	2	8.76	21	7.22	23

Таблица 2. Сравнительная характеристика алгоритмов обучения при фиксированном максимальном числе циклов обучения

Алгоритм обратного распространения ошибки	Фиксированное максимальное число циклов обучения					
	12			100		
	Время обучения, с	Достигнутая точность	Затраченное число циклов обучения	Время обучения, с	Достигнутая точность	Затраченное число циклов обучения
стандартный	3.6	0.008	12	8.27	$4.81 \cdot 10^{-8}$	100
с параметрами сигмоидной функции	2.8	0.001	10	5.03	$2.11 \cdot 10^{-4}$	80
с синусоидальной функцией	2.5	0.001	10	4.21	$3.22 \cdot 10^{-5}$	80
нормировки исходных данных	2.6	0.002	10	3.82	$4.25 \cdot 10^{-11}$	94

Результаты экспериментов свидетельствуют о работоспособности предложенных алгоритмов, о его больших возможностях и позволяют рекомендовать данные алгоритмы обучения нейронных сетей для внедрения и использования на практике.

В главе 4 предлагаются модели нейронных сетей, на базе двунаправленной ассоциативной памяти, для обучения которых применяются разработанные методы обучения нейронных сетей.

Предложена модифицированная модель ассоциативной памяти (рис. 11) на базе двунаправленной ассоциативной памяти, которая имеет наибольшую информационную емкость памяти.

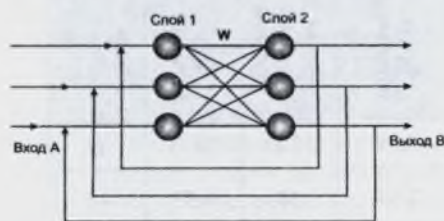


Рис. 11. Модифицированная модель ассоциативной памяти

Для оценки устойчивости предложенной модели создается ассоциативная память, основанная на трехслойную нейронную сеть со структурой 6-8-6. Результаты компьютерного моделирования (Рис. 12) показывают, что обучающая скорость предложенной модели намного быстрее, т.е. в среднем в 1,5 раза быстрее, и кроме того, модель оптически реализуема, т.е. предложенная модель имеет значительную роль при реализации оптических нейронных сетей.

На основе разработанной модели двунаправленной ассоциативной памяти предложена система оптической ассоциативной памяти, структурная схема которой показана на рис. 13. Разработанная система состоит из модулятора пространственного света (ПМС), которая используется как входное устройство, массива линз для реализации массивных межсоединений нейронов, оптической маски, которая служит как матрица взаимосвязи (весовых коэффициентов), ССД камеры.

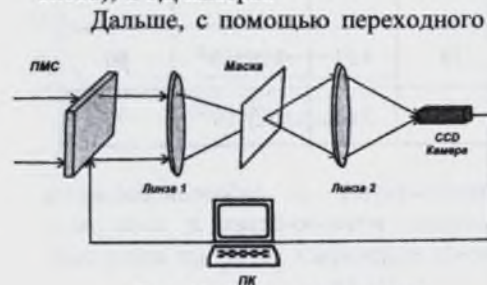


Рис. 13. Структурная схема оптической ассоциативной памяти

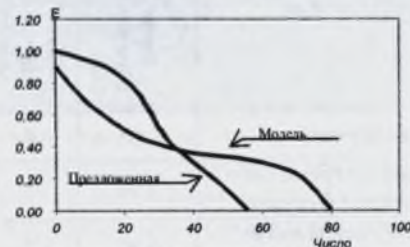


Рис. 12. Результаты тестирования

Дальше, с помощью переходного электронного устройства сопряжения макет подключается к персональному компьютеру, задача которого отображение, обработка и нормализация сигнала, затем обработанный сигнал снова подается в ПМС для следующей обработки.

Данная система используется для классификации символов. В эксперименте были использованы 36 букв алфавита, состоящая из 6x7



Рис. 14. Буква «А» с 30%-ным шумом

пикселей. Буква подвергается 30%-ному случайному шуму (см. рис.14.) и после преобразования подается на вход ПМС. Экспериментально подтверждены эффективность работы предложенной модели по скорости распознавания образов, что распознает 85 символов за секунду.

Результаты теоретических и экспериментальных исследований показывают, что предложенная модель нейронной сети позволяет легко и точно определить оптимальные обучающие условия, имеет большую скорость сходимости и видно простота оптической реализуемости модели.

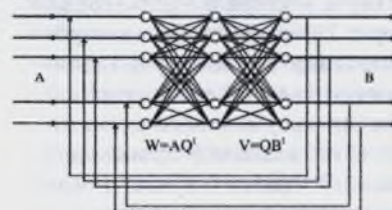


Рис. 15. Схема модели модифицированной ортогональной

Предложена модифицированная ортогональная ассоциативная память (рис. 15), которая имеет тесные связи с другими различными моделями нейронной сети с функциональной точки зрения и является специальным видом гетероассоциативной памяти. На основании теории ортогональной ассоциативной памяти, построим синептические матрицы $W=A \cdot Q^t$ и $V=Q \cdot B^t$. Однако, линейные зависимости создают промежуточный вектор:

$$\bar{q} = a_i^t \cdot W = q_i + \sum_{j=1}^q \cos(a_i^t, a_j) q_j, \quad (7)$$

где \bar{q} - компонента искажения q_i , так как используется итеративный фильтр f^* в скрытом слое.

В модифицированной модели ортогональной ассоциативной памяти уменьшение влияния шумов \bar{q} достигается за счет использования функции активации следующего вида:

$$f(x) = \frac{1}{1 + e^{-x}} - \left(1 - \frac{2}{q}\right), \quad (8)$$

где, q - количество обучаемых изображений.

Функция ошибки, ассоциирующаяся с входным изображением a_i , которая является промежуточным ответом \bar{q} , вычисляется следующим образом:

$$E_i = \begin{cases} 0, & \text{при } \bar{q}_i \geq 0 \\ \frac{1}{2} \sum_{j=1}^q |q_j - \bar{q}_j|, & \text{при } \bar{q}_i < 0 \end{cases} \quad (9)$$

Если ошибка приравняется нулю синаптические веса не изменяются, в противном случае мы применяем правило:

$$\epsilon = \frac{(q_i - \bar{q}_i) \left(1 - \bar{q}_i + \frac{2}{q} \right) \left(\bar{q}_i + \frac{2}{q} \right)}{2q} \quad (10)$$

Для оценки эффективности работы предложенной модели проведен эксперимент, где предложенная сеть обучается ассоциативно восстанавливать изображение. На рис. 16 (а) приведен тестовый набор из трех различных образов (луна, флаг, лицо), связанных с соответствующими надписями. Обучающие данные, составляющие векторы a и b , формировались на базе пиксельных карт, представляющих упрощенный образ изображения (вектор a) и надпись (вектор b). Размерность вектора a равнялась 288 (16x18), а вектора b – 200 (25x8).



а) б)
Рис. 16. Набор для тестирования

На рис. 16 (б) приведен набор с 20% шумом для тестирования сети. После последовательных циклов функционирования модифицированной ортогональной ассоциативной памяти на примере распознавания зашумленного схематического образа изображений и их названий происходит полное распознавание образов. В нашем случае, после трехкратного прохождения сигналов через модифицированную сеть ортогональной ассоциативной памяти произошло безошибочное распознавание как образов, соответствующих изображениям (векторов a), так и связанных с ними надписей (векторов b). На функционирование системы огромное влияние оказывает использование метода обратного распространения ошибки, который в значительной степени зависит от степени искажения образов. Для набора образов, исследовавшихся в описываемом численном эксперименте, обученная при 20%-ном шуме (20% пикселей находились в искаженных состояниях) сеть так же обеспечивала безошибочное распознавание образов, искажение которых достигало 50%.

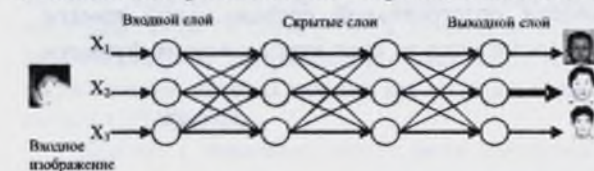
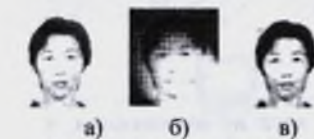


Рис. 17. Структурная схема сети для классификации изображений



а) б) в)
Рис. 18. Тестовый набор изображений

Для оценки помехоустойчивости предложенной системы рассматривается задача распознавания изображений, т.е. изображений сотрудников нашей лаборатории (рис.17). Изображение имеет размер 512x512 пикселей, разделенные на 25-элементные кадры (5x5), на рис.18 представлен тестовый набор изображений: оригинальное изображение (а), искаженное изображение, представленное в виде пикселей (б) и изображение после распознавания (в). Отмечено 96% распознавание на базе данных. Точность распо-

знания составила в среднем 90%, что увеличилась в 1,5 раза. Полученные результаты показали чрезвычайно высокую помехоустойчивость, т.е. даже в условиях высокого внутреннего шума, система способна устойчиво работать и адекватно реагировать на соответствующие входные параметры.

Используемые оптические компоненты с учетом устойчивости нейронных сетей к нарушениям их внутренней структуры значительно более устойчивы к внешним помехам.

ВЫВОДЫ И ЗАКЛЮЧЕНИЕ

В диссертации приведено теоретическое обобщение и решение проблемы разработки новых высокоэффективных методов обучения нейронных сетей и на их основе создание помехоустойчивых оптических систем обработки информации.

В работе получены следующие основные результаты:

1. Предложен алгоритм, позволяющий ускорить сходимость алгоритмов обучения нейронных сетей в случае попадания сети в локальные минимумы. Время обучения нейронных сетей с использованием алгоритма вывода сети из локального минимума сокращалось на 10-30%.

2. Усовершенствован алгоритм обратного распространения ошибки для ускоренного обучения нейронных сетей за счет модифицированных функций активации: сигмоидной функции с различными параметрами, которые позволяют увеличить скорость обучения в 2 раза и синусоидальной функцией с дополнительными фазовыми параметрами, использование которой существенно расширяет возможности оптической реализации нейронной сети. 3. Предложен новый подход нормировки данных с использованием функции активации, который позволяет повысить достоверность, надежность и эффективность нормировки путем приближения распределения данных к равномерному. Результаты экспериментов подтвердили эффективность предложенного способа нормировки, позволившего снизить ошибку обучения сети в 3-4 раза относительно данных, нормировка которых не проводилась.

4. Разработана модифицированная модель двунаправленной ассоциативной памяти, у которой обучающая скорость в 1,5 раза быстрее и на её принципах разработана оптическая система. Результаты экспериментов показали чрезвычайно высокую помехоустойчивость системы, т.е. даже в условиях высокого внутреннего шума, система способна устойчиво работать и адекватно реагировать на соответствующие входные параметры.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ
ДИССЕРТАЦИИ

1. Jumaliev K.M., Akkoziev I.A., Jeenbekov A.A., Sarybaeva A.A., The record and readout image from holographic associative memory // II International Conference on Optical information processing, June 17-21, 1996. - St-Petersburg, Russia. -pp. 140-143
2. Jumaliev K.M., Jeenbekov A.A., Sarybaeva A.A., The method of record and readout 2-D image to HAM // International Conference on Holography and OIP, Nanjing, China. August 26-28, - 1996. -pp. 120-123
3. Акаев А.А., Жээнбеков А.А., Сарыбаева А.А. Возможности применения модели ортогональной проекции в ассоциативной памяти // Труды международного семинара «Голография и оптическая обработка информации», 2-4 сентября 1997г., - Бишкек. - С. 48-51
4. Жумалиев К.М., Жээнбеков А.А., Сарыбаева А.А. Обработка больших массивов данных в голографической ассоциативной памяти // Труды международного семинара «Голография и оптическая обработка информации», 2-4 сентября 1997г., - Бишкек. - С. 52-57
5. Акаев А.А., Жээнбеков А.А., Сарыбаева А.А. Особенности применения модели ортогональной проекции в ассоциативной памяти // Наука и новые технологии, №1, 1998г., - Бишкек. -С. 7-12
6. Jeenbekov A.A., Sarybaeva A.A. Research the conditions of convergence of back-propagation learning algorithm // Proceeding of SPIE, Vol. 4148, 1999. - pp. 12-18.
7. Жумалиев К.М., Жээнбеков А.А., Сарыбаева А.А. Алгоритм последовательной выборки из ассоциативной памяти // Конференция, посвященная 200-летию юбилею А.С. Пушкина в Кыргызстане, КРСУ. 1999. - Бишкек. -С.34-35
8. Сарыбаева А.А. Модифицированная теория для обработки информации в ассоциативной памяти // Конференция, посвященная 200-летию юбилею А.С. Пушкина в Кыргызстане, КРСУ, - Бишкек, 1999. -С. 36
9. Жээнбеков А.А., Сарыбаева А.А. Усовершенствованный алгоритм считывания информации из ассоциативной памяти // Сборник научных трудов КРСУ, отд. Физика, Выпуск 3. , 2000г. - С. 123-126
10. Жээнбеков А.А., Сарыбаева А.А. Исследование влияния синусоидальной пороговой функции на сходимость алгоритма обратного распространения ошибки для обучения нейронных сетей // Наука и новые технологии, №1, 2000 г., - Бишкек. - С. 8-11.
11. Жээнбеков А.А., Сарыбаева А.А. Поиск оптимальных параметров метода обучения нейронных сетей // Материалы международной конференции «Телекоммуникационные и информационные технологии. Состояние и проблемы развития», 6-11 июля 2001 г., - Бишкек. - С.364-368
12. Жээнбеков А.А., Сарыбаева А.А. Оптимизация работы ассоциативной памяти на основе нейронных ансамблей // Материалы международной конференции «Телекоммуникационные и информационные технологии. Состояние и проблемы развития», 6-11 июля 2001 г., - Бишкек. - С.369-373
13. Жээнбеков А.А., Сарыбаева А.А. Оптимальные условия для сходимости алгоритма обратного распространения ошибки // Наука и новые технологии, №1, 2001 г., - Бишкек. - С. 97-100
14. Жээнбеков А.А., Аккозиев И.А., Сарыбаева А.А. Оптическая реализация ассоциативной памяти // Труды международного семинара «Голография и оптическая обработка информации», 5-7 сентября 2001 г., - Бишкек. - С.57-60
15. Жээнбеков А.А., Сарыбаева А.А. Влияние активационных функций на скорость сходимости алгоритма обратного распространения ошибки // Труды международного семинара «Голография и оптическая обработка информации», 5-7 сентября 2001 г., - Бишкек. - С.54-56
16. Жээнбеков А.А., Сарыбаева А.А. Распознавание изображений с помощью аппарата нейронных сетей // Труды международной конференции «Развитие информационно-коммуникационных технологий в информационном обществе: состояние и перспективы», 28-29 октября 2004 г., - Бишкек. - С.83-85
17. Сарыбаева А.А. Алгоритм обратного распространения ошибки с модифицированной функцией ошибки // Труды международной конференции «Развитие информационно-коммуникационных технологий в информационном обществе: состояние и перспективы», 28-29 октября 2004 г., - Бишкек. -С.90-93
18. Сагымбаев А.А., Сарыбаева А.А. Процесс ортогонализации по грамму-Шмидту в голографической ассоциативной памяти // Материалы научной конференции КРСУ (Посвящается дням Славянской письменности и культуры), 15-16 мая 2005г., - Бишкек. -С.61
19. Аккозиев И.А., Жээнбеков А.А., Сарыбаева А.А. Метод оценки ошибки для ускорения обучения нейронных сетей // Наука и новые технологии. №2, 2006г., - Бишкек. - С.8-10
20. Сарыбаева А.А. Алгоритм для ускорения сходимости в случае попадания сети в локальные минимумы // Наука и новые технологии. №2, 2006г., -Бишкек. - С.155-158
21. Жээнбеков А.А., Сагымбаев А.А., Аккозиев И.А., Сарыбаева А.А. Метод оценки ошибки для уменьшения времени обучения нейронных сетей // Вестник КРСУ», Том 7, №8, 2007г., - Бишкек. - С. 60-64.
22. Жээнбеков А.А., Сагымбаев А.А., Сарыбаева А.А. Оптические нейросетевые системы для распознавания изображений // Известия ВУЗов, №1-2, 2007г., - Бишкек. - С.3-5.