

Р. Х. Садыхов, М. Е. Ваткин (vatkin@tut.by)

Алгоритм обучения нейронной сети «неокогнитрон» для распознавания рукописных символов.

Введение. В настоящий момент механизм распознавания визуальных образов, осуществляемых человеческим мозгом, а в частности, такие его особенности, как инвариантность к деформациям образов, изучен недостаточно хорошо. Разными учеными предпринимались попытки моделирования данных процессов [1 – 3], но эти модели не являются инвариантными при решении задач распознавания образов с деформированной формой. Начиная с 1980 г. К. Фукушима опубликовал ряд статей [5 – 8], описывающих нейронную модель называемую "неокогнитрон", которая позволяет устранить эти недостатки. Настоящая работа является продолжением исследований в этой области. В ней описываются результаты моделирования неокогнитрона с оптимизированными по времени исполнения и простоте описания алгоритмами обучения и функционирования сети, а также предлагается новый подход к формированию обучающих образов и связей между слоями сети.

Структура нейронной сети. Общая иерархическая структура сети представлена на рис 1, на котором каждый прямоугольник является 2-мерным массивом нейронов. Первый уровень сети представляет собой слой рецепторных нейронов, которые непосредственно реагируют на входные образы. Каждый последующий уровень имеет слои, состоящие из *S*-нейронов и *C*-нейронов. Последние введены в сеть для того чтобы обеспечить инвариантность к ошибке сдвига локальных признаков, выделяемых данным нейроном. Последний *C*-слой генерирует результаты распознавания.

S- и *C*-нейроны из *l*-го уровня обозначаются соответственно как U_{Sl} и U_{Cl} . Каждый *S*- и *C*-слой нейронов разделен на подгруппы, называемые плоскостями в соответствии с признаками образа которые они выделяют, а каждая плоскость нейронов организована в 2-мерный массив. На рисунке прямоугольники, обведенные жирной линией, обозначают отдельную нейронную плоскость, а большие прямоугольники, объединяющие несколько плоскостей и представленные тонкой линией, обозначают либо *S*-, либо *C*-слои.

В оригинальной модели неокогнитрона применяются также тормозящие нейроны, но в предложенной модели они не используются для упрощения математической части, экономии машинного времени и занимаемого объема памяти. Связи, сходящиеся к нейронам из одной плоскости от предыдущих слоев, являются идентичными, хотя их рецепторные поля отличаются параллельным сдвигом относительно друг друга в соответствии с положением данного нейрона в нейронной плоскости и у каждого такого нейрона имеет место одинаковое распределение весовых коэффициентов этих связей.

Количество нейронов в каждом слое подбирается таким образом, чтобы уменьшаться по мере приближения к выходному слою. Нейроны из последующих слоев имеют рецепторные поля большего размера, поэтому необходимо меньше нейронов для того, чтобы охватить весь входной образ. На рис. 2 изображены продольный разрез сети и принцип определения размера рецепторного поля во входном слое сети для нейрона из

U_{S2} -слоя. На рисунке этот нейрон выделен черным и здесь же представлен принцип объединения нейронов в сеть.

Связи, ведущие к S -нейрону, имеют изменяемые весовые коэффициенты, получаемые в ходе обучения сети. В результате обучения S -нейроны активизируются только тогда, когда в его рецепторном поле оказывается именно тот признак образа, на который обучали данный нейрон. Обычно такие элементы образа (признаки), как линии различной ориентации, выделяются на входных уровнях сети, а на выходных уровнях определяются более глобальные признаки (целый образ либо его крупная часть). S -нейроны введены в сеть для того, чтобы уменьшить ошибку позиционирования. Каждый S -нейрон получает в качестве входа сигналы, поступающие от группы S -нейронов, которые выделяют одинаковый признак образа, но с небольшими смещениями в положении рецепторного поля. Функция активации S -нейрона такова, что S -нейрон переходит в активное состояние тогда, когда хотя бы один из этих S -нейронов находится в активном состоянии, т.е. S -нейрон выполняет функцию нечеткой логики. Таким образом S -нейрон распознает выделяемый элемент образа, даже если он немного сдвинут в своем положении, что обеспечивает инвариантность к сдвигам.

Инвариантность распознавания сетью деформированных в форме образов достигается за счет многоуровневой структуры сети, в которой допускается на каждом слое некоторый параллельный сдвиг при распознавании элементов образа. Следовательно постепенное накопление позиционной инвариантности приводит к инвариантности деформации формы, причем важным фактором является именно постепенное накопление инвариантности сдвига на каждом слое, а не за счет только одного слоя.

Посредством описанного процесса мелкие признаки образа, выделяемые на начальных уровнях сети, постепенно объединяются в крупные признаки, и на выходном уровне каждый отдельный S -нейрон отвечает за распознавание определенной категории входных образов.

Оптимальный размер сети определяется в зависимости от параметров обучающей выборки. Хотя достаточно сложно найти точные критерии выбора размера сети, но можно предложить следующие рекомендации. Если сложность распознаваемого образа велика, т.е. он состоит из большого количества локальных признаков, то объем связей между нейронами должен быть малым, а общее количество уровней сети должно быть большим.

Величина Al в (2) обозначает объем возбуждающих входных связей S -нейрона и соответствует размеру его рецепторного поля в предыдущей плоскости. Величина Al должна определяться из соображений плотности локальных признаков образа. Чем сложнее образ, т.е. содержит больше мелких по размеру признаков, тем пропорционально этому должна быть меньше величина Al . На выбор оптимальной величины Al также влияет уровень инвариантности к деформациям, который должна проявлять сеть в особенности на входных слоях. Величина Al может быть большой, если вероятность деформации локальных признаков образа мала. Когда к сети предъявляется требование нечувствительности к большим искажениям, эта величина должна быть малой.

Величина Dl в (3) определяет объем входных фиксированных возбуждающих связей S -нейрона, поступающих от отдельной предыдущей плоскости, и проявляет аналогичные тенденции, что и Al . Если плотность признаков образа велика, то величина Dl должна быть малой. Когда величина Dl велика, это означает, что на данном уровне сети может быть

обработана большая деформация образа и что сеть может быть нечувствительной к данной деформации образа благодаря меньшему количеству уровней.

Структура представленной модели неокогнитрона для распознавания рукописных цифр и букв насчитывает три уровня. Последний S -слой содержит 36 нейронных плоскостей, из которых каждая содержит только один нейрон соответствующий отдельному классу символов. Символы делятся на следующие категории: заглавные символы русского алфавита, исключая символы Ё, Ъ, Щ, Ь всего 29 букв и 7 цифр. Буква О и цифра 0, буква Ч и цифра 4, а также буква З и цифра 3 имеют совмещенные классы, в силу их одинакового написания. Некоторые стили написания цифры 6 идентичны со стилем написания буквы Ъ, что также позволяет объединить их в один класс.

Алгоритм обучения. В процессе обучения сети происходит настройка весовых коэффициентов входных связей S -нейронов. Обучение производится последовательно от входных слоев сети к выходным. На рис 2 изображен обобщенный алгоритм обучения сети. Все слои обучаются идентичным методом. Допустим, что обучение производится для некоторого U_{Sl} -слоя. Тогда для этого слоя выбираются очередная плоскость k и нейрон в этой плоскости $U_{Sl}(k)$, который будет обучаться. Данный нейрон называется «корневым» и он указывает позицию центра рецепторного поля. Затем на входном слое сети в центр рецепторного поля загружается обучающий образ. Далее происходит последовательная активация нейронов всех ранее обученных слоев вплоть до слоя, который непосредственно передает сигнал обучаемому слою сети. После этого происходит обучение выбранного нейрона. Поскольку для нейронов из одной плоскости имеем одинаковое распределение весовых коэффициентов, то логичным является обучение только одного нейрона из этой плоскости, а весовым коэффициентам остальных нейронов следует присвоить идентичные значения.

В данной работе анализировались алгоритмы обучения и функционирования сети с целью упрощения их математической модели, а как следствие, оптимизация по времени их исполнения в терминах количества элементарных математических действий. В связи с этим проводился анализ литературы [10 – 12], где описываются алгоритмы обучения нейронов сети Кохонена и нейронов на основе радиальной базисной функции. Как результат предлагается перейти от сложной оригинальной структуры слоя S -нейронов, который использует тормозящие связи и нейроны, требующие дополнительной памяти и математических операций, к структуре S -нейрона, изображенной на рис. 3. Фактически для обучения S -нейрона предлагается правило обучения нейронов конкурентной сети Кохонена, а для активации – правило активации нейронов на основе радиальной базисной функции. В результате применения предложенного подхода удалось сократить объем оперативной памяти, необходимый для хранения весовых коэффициентов, а также количество операций умножения в два раза, за счет отказа от тормозящих нейронов и связей.

Таким образом, функция обучения S -нейрона будет соответствовать соотношению вычисления динамического среднего для каждого весового коэффициента

$$w(t+1) = w(t) + \frac{1}{t+1} \cdot (x(t+1) - w(t)), \quad (1)$$

а функция активации S -нейрона будет соответствовать радиальной базисной функции (рис. 3)

$$U_{Sl}(n, k) = \exp \left[\sqrt{\frac{\sum_{p=1}^{P_{Cl-1}} \sum_{v \in Al} (U_{Cl-1}(n+v, p) - w(v, k))^2}{N}} \right], \quad (2)$$

где l – порядковый номер слоя; k – номер обучаемой плоскости; n – двумерный индекс нейрона в k -й плоскости; w – весовой коэффициент связи; v – двумерное смещение входной связи в пучке связей Al ; Al – двумерная величина характеризующая объем входных связей S -нейрона идущих от одной плоскости из слоя U_{Cl} ; N , p – общее количество входных связей и порядковый номер плоскости, соединенной с обучаемым нейроном соответственно.

C -нейроны активизируются в соответствии с функцией ИЛИ нечеткой логики

$$U_{Cl}(n, k) = \max(U_{Sl-1}(n+v, p)), \quad \forall v \in Dl, \forall p \in P, \quad (3)$$

где p , P – порядковый номер плоскости и множество плоскостей из слоя U_{Sl} , которые имеет связь с активируемым нейроном, соответственно; Dl – двумерная величина, характеризующая объем пучка входных связей C -нейрона, идущих от одной плоскости из слоя U_{Sl} ; v – двумерный индекс связи внутри пучка связей Dl .

На рис 2 изображен принцип расчета рецепторного поля нейрона, который называют "корневым" нейроном (выделен черным цветом). А принцип вычисления рецепторного поля похож на принцип роста кристалла от точки кристаллизации, которой является «корневой» нейрон.

Стратегия выбора обучающих образов сети. Каждый образ предоставляемый сети для распознавания рассматривается в данной статье в виде совокупности элементов образа – его признаков. Все признаки образа можно разбить на две группы: неделимые и составные. Неделимые признаки рассматривают как некие цельные, структурные элементы образа, составные признаки являются комбинацией неделимых признаков. Так как детализацию образа можно проводить до бесконечности, то вводится критерий неделимости, который требует от неделимой признака сохранения свойства структуры. Например, точка не несет в себе свойства структуры, а две точки, расположенные рядом, уже являются отрезком линии, который прослеживает такое свойство структуры, как определенная ориентация. С этой точки зрения элементы образов, из которых состоят буквы русского алфавита можно разбить на такие неделимые элементы, как отрезки линий. Далее определяется такие свойства элементов образа, как «точечность» и «протяженность». Точечность элемента образа говорит о том, что свойства данного элемента скорее сконцентрированы в одной точке, чем распределены в пространстве. Например, пересечение двух прямых является точечным элементом, а линия – протяженным элементом, так как его свойства распределены на промежутке между точками начала и конца линии.

Как правило, от входных слоев нейронной сети к выходным слоям признаки, выделяемые плоскостями этих слоев прослеживают следующие правило. На входных слоях выделяют неделимые элементы образа, обладающие свойством протяженности, такие как отрезки линий различной ориентации, а по мере приближения к выходным - составные элементы образа, обладающие свойством точечности. Этот подход обоснован следующим. Допустим, что в средних слоях наряду с точечными элементами выделяются также и протяженные элементы, как линии. Тогда в плоскостях данных слоев, которые

выделяют точечные элементы, будет активизирован всего один или некоторая небольшая группа нейронов, а в плоскостях, выделяющих протяженные элементы, будет активна целая группа нейронов на протяжении длины всей линии. Таким образом, если выходной сигнал с последней плоскости будет оцениваться по тем же критериям, как и выходной сигнал с плоскости, выделяющей точечные признаки, то влияние этого сигнала будет неоправданно большим, что приведет к неадекватной оценке изображения. В связи с этим на входных слоях рекомендуется применение протяженных признаков, а по мере приближения к выходным слоям – точечных.

В оригинальных статьях К. Фукушима [7 – 9] изображение символа, поступающего на вход сети, вводится специальным образом, то есть символ не сканируется, а вводится с помощью манипулятора «мышь», поэтому эти изображения имели фиксированную ширину линий. Поскольку в данной работе решается задача распознавания рукописных заглавных символов русского алфавита, написанных разными людьми, а изображения сканируются с бумажного носителя, то толщина линии таких изображений колеблется даже в пределах написания одного символа. Это, усложняет процесс анализа изображения и делает невозможным применение подхода, который использован в [9]. В связи с этим авторами предлагается производить оценку не самих линий, а границу перепада яркости между светлым фоном и темными точками символа. Данная задача решается с помощью создания специальных обучающих образов для входного слоя U_{S1} , которые будут выделять именно границы образа. Так как выходная информация с входного слоя связана только с границами перепадов яркости, это позволяет последующим слоям учитывать только с эти признаки.

Обучающие образы для слоя U_{S1} . Слой U_{S1} обучается так, чтобы выделять линейные отрезки границ образов, различной ориентации. Поскольку рецепторное поле во входном слое для нейронов из слоя U_{S1} имеет размер 2×2 , то размер обучающего образа тоже составляет 2×2 точки. Подобная связь также прослеживается для всех обучаемых слоев.

На рис. 4 изображены 8 обучающих образов для 8 плоскостей слоя U_{S1} . Образы, объединенные скобкой, обучают такие S -плоскости, нейроны которых имеют общие корневые нейроны из последующей C -плоскости, как показано на рис 5. Это необходимо для того, чтобы C -нейрон с подобными связями имел максимальную активность независимо от того, какую пару границ (левая-правая для вертикальных границ, верхняя-нижняя для горизонтальных границ и т.д.) выделяют S -нейроны из предыдущих плоскостей, но учитывал только ориентацию этих границ. Данный принцип не отображен на рисунке 1, с целью упрощения понимания и по причине того, что это является нововведением в архитектуре неокогнитрона. Таким образом, в слое U_{C1} будет содержаться не 8 плоскостей, а 4 в соответствии с выделяемыми направлениями.

Величина A_1 , определяющая объем пучка связей ведущих к отдельной предыдущей плоскости для слоя U_{S1} , совпадает с его рецепторным полем. Величина D_1 , определяющая рецепторное поле в предыдущем слое U_{S1} для нейронов из слоя U_{C1} , составляет размер 2×2 , что также можно наблюдать на рис. 2.

Обучающие образы для слоя U_{S2} . Слой U_{S2} обучался так, чтобы выделять элементы изображения, которые являются различными сочетаниями элементов изображения, выделяемых на слое U_{S1} (окончания линий, острые тупые и прямые углы различной ориентации). Обучающие образы для слоя U_{S2} представлены на рис 6. Их размерность

составляет 5×5 точек. Образы, объединенные скобкой, обучают одну и ту же S -плоскость. Величина A_2 равна 3×3 , величина $D_2 - 2 \times 2$.

Обучающие образы для слоя U_{S3} . Слой U_{S3} обучается так, чтобы производить уже непосредственно классификацию всего изображения. Размер рецепторного поля S -нейронов составляет 26×26 , а размер обучающих образов – 26×19 точек. Примеры обучающих образов для слоя U_{S3} показаны на рис. 7. Величина A_2 равна 21×21 , величина D_2 равна 3×3 . Образы, объединенные сплошной скобкой, обучают одну S -плоскость. Образы, объединенные штриховой скобкой, обучают разные S -плоскости, но имеющие исходящие связи, как показано на рис 5.

Тестирование. Тестовая выборка была составлена из набора символов русского алфавита, исключая символы Й, Ё, Щ, Ъ 29 букв и 10 арабских цифр. Объем вариантов написания каждого символа составил 100 экземпляров, общий объем тестовой выборки – 3900 символов. Результаты тестирования представлены в таблице. Тестирование показало в среднем уровень распознавания символов, равный 91,5%.

Summary

In the paper the experimental simulation results of a neural network "neocognitron" are presented. The algorithm of network training optimized on execution time and complexity of the description are proposed. Also the new approach to configuration of teaching images and network interlayer links is described.

Литература

1. Marko H. and Giebel H. // Nachrichtentechnische Zeitschrift. 1970. Vol. 23. P. 455–459.
2. Marko H. // IEEE Trans. Syst., Man, Cybern. 1974. Vol. SMC-4. P. 34–39.
3. Fukushima K. // Biol. Cybern. 1975. Vol. 20. P. 121–136.
4. LeCun Y. et al. // Neural Computat. 1989. Vol.1. P. 541–551.
5. LeCun Y. et al. // IEEE Communications. 1989. Vol.27. N 11. P. 41–46.
6. Fukushima K. // Biol. Cybern. 1980. Vol. 36. P. 193–202.
7. Fukushima K. and Miyake S. // Pattern Recognition. 1982. Vol. 15. P. 455–469.
8. Fukushima K. Miyake S. and Ito T. // IEEE Trans. Syst., Man, Cybern. 1983. Vol.SMC-13. P. 826–834.
9. Fukushima K. and Wake N. // IEEE Trans. On Neural Networks. 1991. Vol. 2. N 3, P. 355–365.
10. Kohonen T. Self-Organization Maps. Springer, 1995.
11. Powell M.J.D. In: J.C. Mason and M.G. Cox (Eds.). Algorithm for approximation. Oxford: Clarendon Press, 1987. P.143–167
12. Foor W. // Applied optics. 1995. Vol. 34. N 32. P. 7545–7555

*Институт технической
Кибернетики НАН Беларуси*

Поступила в редакцию

УДК 007.001.362; 681.327.12.001.362

Садыхов Р.Х., Ваткин М.Е. **Алгоритм обучения нейронной сети «неокогнитрон» для распознавания рукописных символов.** // Весці НАН Беларусі. Сер. фіз.-тэхн. навук. 2002. N 3. С. 1

Выполнено экспериментальное моделирование нейронной сети «неокогнитрон». Представлены оптимизированные по времени исполнения и простоте описания алгоритмы обучения и функционирования сети, а также предлагается новый подход к формированию обучающих образов и связей между слоями сети.

Табл. 1. Ил. 7. Библиогр. – 12 назв.

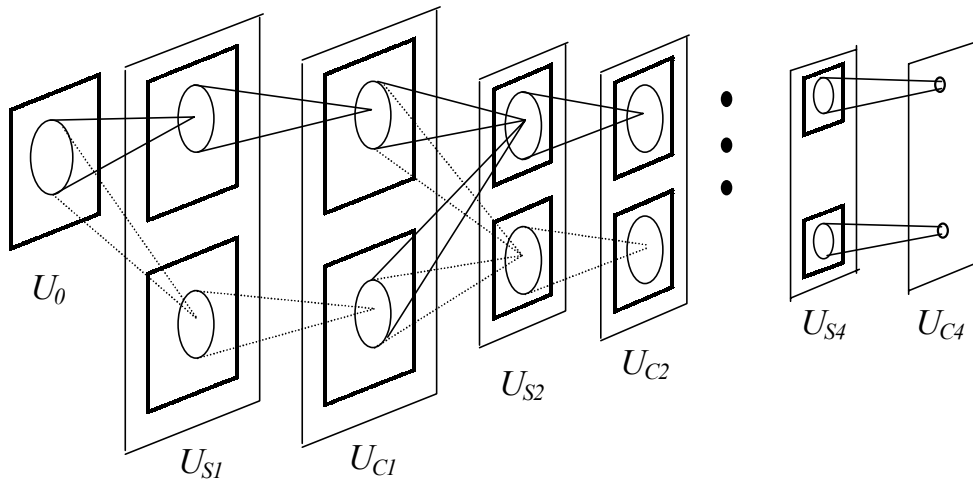


Рис. 1

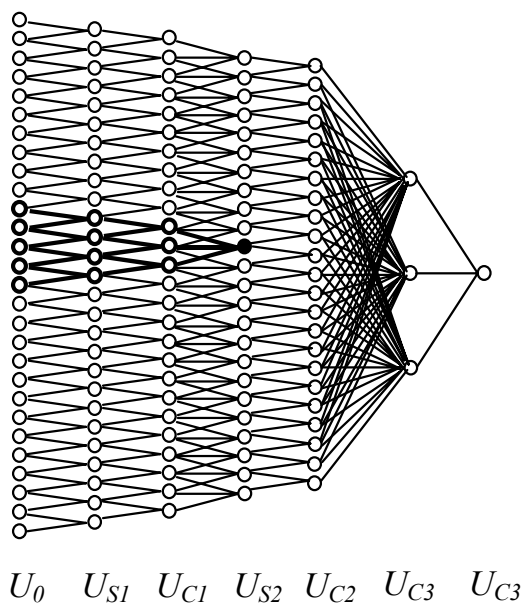


Рис. 2

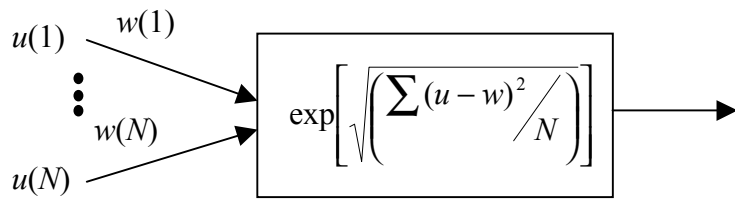


Рис. 3

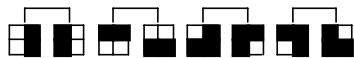


Рис. 4

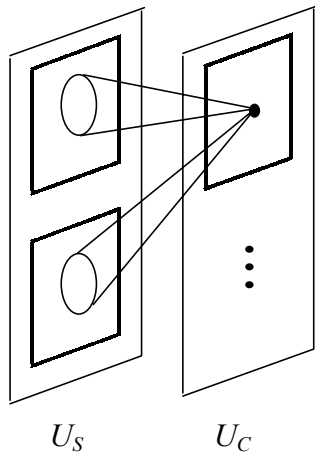


Рис. 5

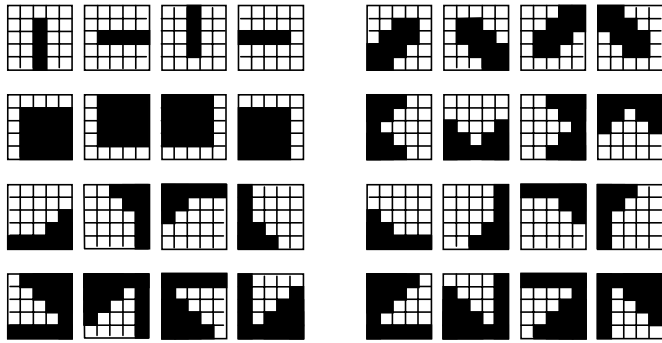





Рис. 6



Рис. 7

Результаты тестирования

Примеры идентификации символов		Процент ошибки по категориям символов
Правильно	Неправильно	
		7,3
		8,2
		12,1
		3,6
		15,8
		9,7
		7,5
		8,9
		11,8
		7,8
		8,2
		3,9
		13,2
		5,3
		6,2
		9,7
		4,1
		3,7
		13,4
		5,8
		14,1

44444444	44	12,7
24244444	4	10,5
44444444	4	12,9
66666666	66	11,7
66666666	6	8,5
33333333	333	4,7
88888888	888	3,1
88888888	8	6,2
11111111	11	3,8
22222222	22	7,8
55555555	353	10,7
66666666	66	11,5
77777777	77	8,6
88888888	888	9,3
99999999	999	6,7

Рис. 1 Структура неокогнитрона

Рис. 2 Структура неокогнитрона в разрезе

Рис. 3 Структура S -нейрона

Рис. 4 Обучающие образы для слоя U_{S1}

Рис. 5 Принцип объединения в слое U_C неидентичных признаков образа, выделяемых в слое U_S .

Рис. 6 Обучающие образы для слоя U_{S2}

Рис. 7 Обучающие образы для слоя U_{S3}

Таблица Результаты тестирования