

Программный эмулятор нейросети типа неокогнитрон для распознавания графических образов

Федяев О.И., Махно Ю.С.

Донецкий национальный технический университет
fedyaev@r5.dgtu.donetsk.ua, ymakhno@gmail.com

Abstract

Fedyayev O., Makhno Y. For recognition of graphic images in the presence of distortions the program model of neocognitron neural network is developed. The used structure of neocognitron allows recognizing images even with its rotations. The training algorithm and feature of neocognitron functioning are considered. Authors note its merits and demerits at recognition of the images liable to distortions.

Введение

Не снижающий интерес к распознающим задачам обусловлен необходимостью автоматизации как функций контроля и управления сложными динамическими объектами в реальном времени, так и образных процессов коммуникации (визуальных, речевых) в интеллектуальных системах. Поэтому до сих пор продолжается поиск и реализация эффективных принципов передачи распознавательной функции человека компьютеризированным системам. Одно из перспективных направлений решения данной проблемы основано на применении нейрокомпьютеров, как наиболее адекватных классу трудно формализуемых задач распознавания образов. В настоящее время предложено большое количество нейросетевых парадигм для решения задач распознавания образов. Значительные затруднения при распознавании вызывают образы, подверженные какому-либо искажению (зашумлению, смещению, повороту, изменению образа в размерах). Эту проблему решают путём выбора соответствующей архитектуры и способа обучения. Анализ работ показывает, что пока не существует такой модели, которая была бы не чувствительна ко всем четырём видам искажений. Проблема достаточно хорошо решена относительно смещённых и зашумлённых образов нейросетями обратного распространения ошибки. Однако, по-прежнему, вызывают затруднение такие виды искажений, как изменение образа в размерах и их поворот.

Перспективу в преодолении этих трудностей связывают с использованием новой нейросетевой парадигмы – моделью неокогнитрона, использующей качественно новую архитектуру и неконтролируемое обучение [1,2,3,6]. В основу архитектуры неокогнитрона положена организация зрительной системы человека. Первая модель неокогнитрона была разработана в 1980 г.

К.Фукушимой [1] и являлась последующим развитием нейросети когнитрон.

В данной работе рассматривается усовершенствованный неокогнитрон (R-неокогнитрон), ориентированный на распознавание развёрнутых образов [2]. С помощью созданной программной модели R-неокогнитрона проведен анализ процесса обучения и распознавания графических образов. Результаты моделирования показали, что отличительной особенностью нейросетей типа неокогнитрон является их высокая скорость обучения и распознавания, а также ошибкоустойчивость при распознавании образов, подверженных всем видам искажений. Благодаря своей архитектуре R-неокогнитрон позволяет распознавать образы, развёрнутые более чем на 30 градусов.

К нерешённым вопросам можно отнести отсутствие точной формулировки и решения задачи оптимизации многочисленных параметров модели неокогнитрона, адекватных природе распознаваемых графических образов.

Модель неокогнитрона

Неокогнитрон является иерархической нейронной сетью, состоящей из достаточно большого числа идущих друг за другом слоёв и имеющей неполные (можно сказать достаточно редкие) связи между слоями [1,2,4]. Модель неокогнитрона была разработана после того, как в зрительной системе были обнаружены узлы, реагирующие на такие элементы, как линии и углы определенной ориентации. На более высоких уровнях узлы реагируют на более сложные и абстрактные образы такие, как окружности, треугольники и прямоугольники. На ещё более высоких уровнях степень абстракции возрастает до тех пор, пока не определяются узлы, реагирующие, например, на лица или другие сложные формы. В общем случае узлы на более высоких уровнях получают вход от группы низкоуровневых узлов и, следовательно, реагируют на более широкую область

визуального поля. Реакции узлов более высокого уровня менее зависят от позиции и более устойчивы к искажениям [3]. Схематический принцип процесса распознавания R-неокогнитроном представлен на рис.1.

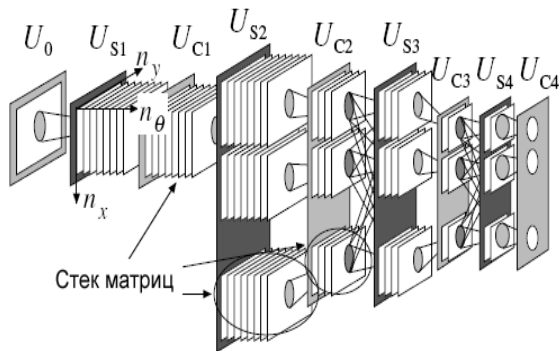


Рисунок 1 – Иерархическая структура R-неокогнитрона

Каждый слой неокогнитрона состоит из стеков матриц. Каждая плоскость в конкретной группе отвечает за определённый угол поворота образа, закреплённого за данной группой. Количество плоскостей зависит от необходимой точности в определении развёрнутых образов. Все матрицы одного слоя имеют одинаковую размерность. От слоя к слою размерность матриц уменьшается либо остается прежней [4]. Элементами матриц являются отдельные нейроны. Все нейроны в каждой матрице реагируют на один и тот же образ (рис.2). То есть, каждый из них настраивается на один специфический входной образ. Каждый нейрон чувствителен к ограниченной области входного образа, называемой его рецептивной областью.

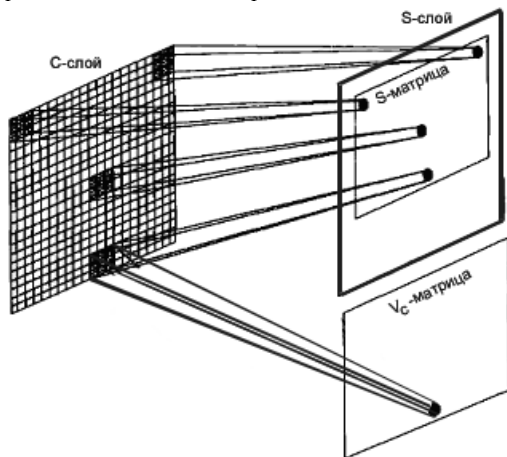


Рисунок 2 – Связи между нейронами S-слоя и нейронами области видимости C-слоя

В неокогнитроне выделяют два типа слоёв: «простые» (S-слой, от слова simpl) и «комплексные» (C-слой, от слова complex), которые состоят соответственно из «простых

матриц» (распознают входную информацию) и «комплексных матриц» (обобщают распознанную информацию). Нейроны простых матриц связаны только с некоторыми нейронами комплексных матриц. В свою очередь нейроны комплексных матриц связаны с нейронами простых матриц следующего слоя.

Входной слой U_0 обычно представляется как нулевой комплексный слой [4]. Концептуально, комплексные матрицы созданы для того, чтобы объединить однотипные результаты распознавания простых матриц. Например, если нейроны первой простой матрицы распознают символ «g», а второй матрицы символ «c», то эти результаты в комплексной матрице следует объединить, т.к. это одна и та же английская буква.

Каждый нейрон как простого, так и комплексного слоя получает сигналы не от всех нейронов, а лишь от некоторых, с которыми он связан. Такие нейроны образуют «область видимости» [4,5]. В отличие от термина «рецептивное поле», термин «область видимости» определяет нейроны предыдущего слоя, с которыми связан рассматриваемый нейрон, а рецептивное поле определяет ту часть входного слоя, которую распознает данный нейрон. Для центрального нейрона каждой матрицы, центральный элемент области видимости располагается в центре матрицы предыдущего слоя (рис.2).

Различают слои, которые могут осуществлять и не выполнять сжатие графической информации. Для слоёв со сжатием характерно то, что размерность его матриц меньше, чем размерность матриц предыдущего слоя. Для сжатия могут использоваться различные методы. В созданной модели для слоёв со сжатием используется следующий метод: центр области видимости нейрона, смещенного на 1 позицию от центрального нейрона, смещается на 2 позиции относительно центра матрицы предыдущего слоя [4].

При определении математической модели неокогнитрона для каждого нейрона простой матрицы, вводится дополнительный тормозящий (или подавляющий) нейрон v . На тормозящий нейрон возлагается функция подавления выходного сигнала нейрона при несоответствии входного образа образу, на распознавание которого настроен нейрон. Подавляющий нейрон связан с нейроном простого слоя и с теми же нейронами комплексного слоя, с которыми связан сам простой нейрон.

При математическом представлении модели введём два дополнительных вектора. Вектор n и вектор k . Вектор n задает положение нейрона в стеке матриц (рис.1):

$$\bar{n} = (n_x, n_y, n_\theta) \quad (1)$$

Вектор k задаёт смещение координат нейрона относительно рассматриваемого нейрона (вектор n) в стеке матриц предыдущего слоя. Он также имеет 3 координаты:

$$\bar{k} = (k_x, k_y, k_\theta) \quad (2)$$

Введем оператор \oplus_M между этими двумя векторами:

$$\left\{ \begin{array}{l} n_x \oplus_M k_x = n_x + k_x \\ n_y \oplus_M k_y = n_y + k_y \\ n_\theta \oplus_M k_\theta = (n_\theta + k_\theta) \bmod M \end{array} \right. \quad (3)$$

Обозначим через параметры T_c и T_s количество матриц в стеках матриц предыдущего слоя. С помощью множества A определим все нейроны, входящие в область видимости рассматриваемого нейрона. Это множество состоит из трёхмерных векторов, каждый из которых задаёт относительное смещение нейрона области видимости от вектора n .

Для расчёта выходного сигнала подавляющего нейрона используется следующая формула:

$$v(\bar{n}) = \sqrt{\sum_{k \in A} t(k) [c(\bar{n} \oplus_{T_c} \bar{k})]^2}, \quad (4)$$

где c – выходные сигналы нейронов предыдущего комплексного слоя; t – весовые коэффициенты от нейронов S -слоя к подавляющему нейрону.

Сигнал, подаваемый на вход простого нейрона, вычисляется по формуле:

$$x(\bar{n}) = \frac{1 + \sum_{k \in A} c(\bar{n} \oplus_{T_c} \bar{k}) \cdot w(\bar{k})}{1 + v(\bar{n}) \cdot w_v} - 1, \quad (5)$$

где w – настраиваемые коэффициенты; v – выходной сигнал подавляющего нейрона (4).

Функция активации нейронов простого слоя имеет вид:

$$S(x) = \begin{cases} x, & x \geq 0 \\ 0, & \text{иначе} \end{cases}, \quad (6)$$

Входной сигнал для нейронов комплексного слоя вычисляется по формуле:

$$y(\bar{n}) = \sum_{k \in A} s(\bar{n} \oplus_{T_s} \bar{k}) u(\bar{k}), \quad (7)$$

где s – выходной сигнал простых нейронов; u – весовые коэффициенты от простых нейронов к комплексному нейрону [4].

Для нейронов комплексных слоёв использовались различные функции активации.

Обучение неокогнитрона

Для обучения неокогнитрона применялась стратегия «обучение без учителя». Настраиваемыми коэффициентами выступали коэффициенты от нейронов комплексного слоя к нейронам простого слоя и коэффициент от подавляющего нейрона к нейрону простого слоя. Коэффициенты от комплексных нейронов к подавляющему нейрону и от простых нейронов к комплексным являются фиксированными. В связи с этим возникает задача определения значений фиксированных коэффициентов. В самом начале предполагается, что значения коэффициентов должны уменьшаться с удалением от центра области видимости.

При настройке весовых коэффициентов использовалась общепринятая формула для расчёта параметра Δ , зависящего от величины входных сигналов [4,5]. В итоге применялись следующие формулы корректировки коэффициентов для нейрона заданного вектором n :

$$\begin{aligned} \Delta w(\bar{k}) &= \alpha \cdot t(\bar{k}) c(\bar{n} \oplus_{T_c} \bar{k}); \\ w^{(m+1)}(\bar{k}) &= w^{(m)}(\bar{k}) + \Delta w(\bar{k}) \end{aligned} \quad (8)$$

Величина Δ для весового коэффициента подавляющего нейрона определялась так, чтобы процесс обучения мог завершиться. Поскольку нет единого мнения по поводу наиболее подходящей формулы для расчёта Δ , то оценивались следующие два возможных способа.

Первый способ основывался на формуле $\Delta w_v = \alpha c(\bar{n} \oplus_{T_c} mid)$ [4]. В этом случае настраиваемый коэффициент зависит только от центрального элемента предыдущего слоя. Это накладывает некоторые ограничения на расположение обучающего образа, то есть образ должен быть расположен так, чтобы нейроном предыдущего слоя с максимальным возбуждением был нейрон, который находится в центре области видимости. Такое ограничение накладывает в силу того, что в противном случае в процессе обучения числитель дроби, используемой при расчёте входа на S -нейрон, будет расти быстрее знаменателя, что приведёт к невозможности обучения.

Второй способ основывался на формуле $\Delta w_v = \alpha \frac{v}{2}$ [5]. В этом случае дельта зависит от выхода тормозящего нейрона, а сам выход тормозящего нейрона зависит от всех нейронов предыдущего слоя, с которыми связан обучаемый нейрон. В программной модели применялся второй способ из-за его универсальности.

Обучение неокогнитрона происходит слой за слоем. При обучении нейронов использовался метод, при котором в каждой матрице обучается только один нейрон (обычно используется центральный нейрон). После завершения обучения значения весовых коэффициентов этого нейрона распространяются на все остальные нейроны в этой матрице [4]. Таким образом, процесс обучения производится достаточно быстро, так как на каждом слое необходимо обучить лишь столько нейронов, сколько матриц в данном слое, а их количество обычно не превосходит 100 [3,4].

Для обучения нейронов первого слоя шаблон обучения записывается на входном слое и после этого начинается обучение центральных нейронов каждой матрицы. После того, как первый слой обучен, переходят к обучению второго слоя. Для этого необходимый образ записывают на входном слое, вычисляют выходы всех нейронов первого простого слоя, первого комплексного слоя и на основе полученных результатов (выходов нейронов первого комплексного слоя) обучают нейроны второго простого слоя. Для третьего и последующих слоёв всё аналогично (вычисление выходов нейронов на втором, третьем и других слоях происходит лишь после записи шаблона обучения) [4,6].

Нейроны последнего слоя должны видеть полностью входной образ, иначе теряется смысл распознавания. Это условие является обязательным, но, как показали исследования, размер рецептивного поля не обязательно должен быть равным размеру входного изображения лишь на последнем слое, это может происходить и раньше.

Анализ процесса распознавания

Для исследования процессов обучения и распознавания была разработана программа-эмулятор нейронной сети R-неокогнитрон.

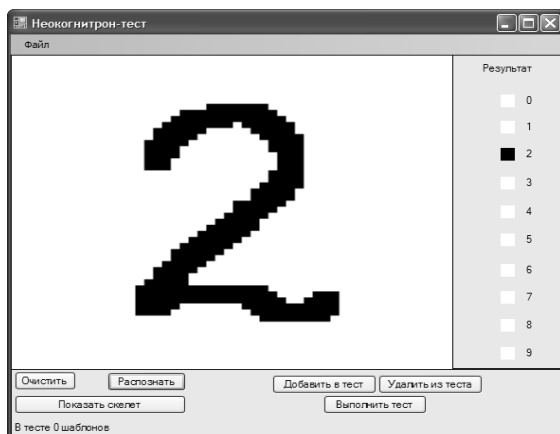


Рисунок 3 – Окно программного эмулятора R-неокогнитрона

Для обучения неокогнитрона на распознавание рукописных арабских цифр использовались шаблоны обучения из открытой базы данных ETL1. Главная форма программы-эмулятора представлена на рис.3.

Анализ процесса распознавания был проведен на классах образов, обеспечивающих проверку чувствительности неокогнитрона к:

- изменению масштаба;
- изменению местоположения;
- присутствию хаотических шумов;
- повороту исходного образа;
- наложению образов.

Трансформации подвергались лишь те образы, которые были правильно распознаны неокогнитроном в оригинальном виде.

При изменении масштаба (рис.4) неокогнитрон показал 92% распознаваемости. Ошибки в распознавании происходили в основном при сильном уменьшении масштаба.

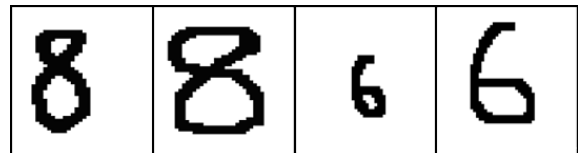


Рисунок 4 – Распознавание при изменении масштаба

Неокогнитрон является нейронной сетью нечувствительной к местоположению образов (рис.5). Это было подтверждено результатами моделирования. Неокогнитрон показал 100% распознаваемость на предложенных образах.

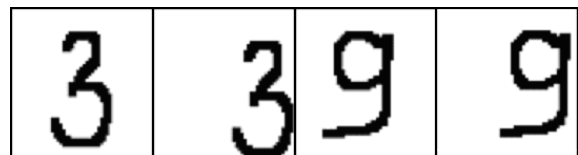


Рисунок 5 – Распознавание при изменении местоположения

Отличные характеристики неокогнитрон показал и при добавлении хаотических шумов (рис.6). Были распознаны правильно 94% образов. Ошибки возникали лишь в случаях заполнения хаотическими шумами более 20% всей рабочей области.

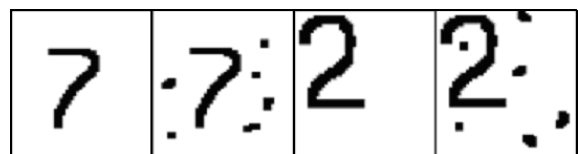


Рисунок 6 - Распознавание при добавлении хаотических шумов

Данная модель неокогнитрона также позволяет распознавать образы, повернутые не более чем на 70 градусов от своего исходного положения (рис.7). При повороте образов неокогнитрон показал результат распознавания, равный 79%.

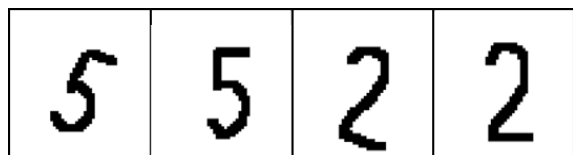


Рисунок 7 - Распознавание при повороте исходного образа

Одним из недостатков неокогнитрона является то, что он неправильно реагирует на образы, которые были получены путём наложения одного образа на другой. Как показано на рис.8, при подаче образа, который был получен из образов цифр 2 и 5, нейронная сеть выдала ответ, обозначающий цифру 8.

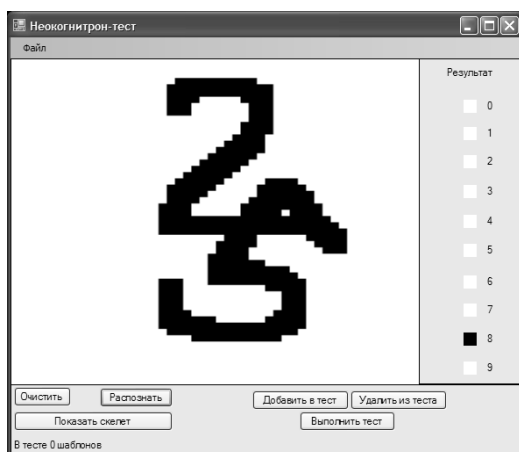


Рисунок 8 - Сложность распознавания при наложении образов

Выводы

Моделирование процессов обучения и распознавания на программной модели R-неокогнитрона показало перспективы распознавания графических образов в условиях различных видов помех, в том числе и при изменении их ориентации. Она обусловлена особой структурой данного типа нейросетей, в определённой степени имитирующей работу зрительной системы человека. Неокогнитрон можно рассматривать как многоуровневую нейросетевую систему индуцирования знаний из обучающего множества графических данных, посредством выделения типовых признаков и их обобщения на последующих уровнях.

Обобщая результаты исследований процесса распознавания R-неокогнитроном

графических образов, подверженных различного рода искажениям, отметим ряд его свойств, благодаря которым неокогнитрон превосходит другие нейросети, ориентированные на распознавание образов:

- малая чувствительность к основным искажениям, которым подвергаются графические изображения (масштабирование, перенос, хаотические шумы);
- обучение неокогнитрона происходит за приемлемое время; способность решать задачу распознавания графических образов за короткое время;
- неокогнитронные нейросети самоорганизующиеся;
- наличие в неокогнитроне достаточного количества параметров для его настройки на распознавание различных типов образов;
- неокогнитрон сам выделяет отличительные особенности во входном наборе образов, которые в дальнейшем используются при идентификации каждого образа независимо от деформации.

Дальнейшая работа будет направлена на оптимизацию параметров архитектуры неокогнитрона. Построение на базе неокогнитрона эффективных систем обработки образной информации позволит решать многие практические задачи в различных предметных областях: распознавание сигналов, идентификация графических изображений, управление с использованием средств визуального контроля за объектами.

Литература

1. Fukushima K. Neocognitron: a self-organising neural network for mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* 36, 1980. – P. 193-202.
2. Федяев О.И., Куликов С.А. Многоуровневые нейронные модели типа неокогнитрон. Архитектура, обучение и распознавание // Труды 8-й Всероссийской конф. «Нейрокомпьютеры и их применение» (НКП-2002). – М.: Век книги, 2002. – С.1046-1052.
3. Уоссермен Ф. Нейрокомпьютерная техника. Теория и практика. – М.: Мир, 1992. – 260 с.
4. Fausett L. *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*. Prentice Hall, New Jersey, 2000.
5. Satoh S., Kuroiwa J., Aso H. and Miyake S. Recognition of Hand-written Patterns by Rotation-invariant Neocognitron – Proc. of ICONIP'98, 1998. – P. 295-299.
6. Федяев О.И., Куликов С.А. Применение многоуровневых нейронных сетей типа неокогнитрон для распознавания символов // Сб. тр. междунар. науч. конф. «Нейросетевые технологии и их применение». – Краматорск: ДГМА, 2003. – С. 247-256.