

УДК 004.93

СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ С НЕПОЛНОЙ МАСКОЙ ВЕСОВ ДЛЯ ЗАДАЧИ РАСПОЗНАВАНИЯ РУКОПИСНЫХ ЦИФР

Тимощенко П.А.

*Днепропетровский национальный университет им. О.Гончара
Кафедра автоматизированных систем обработки информации
e-mail: t_pa@mail.ru*

Анотация

Тимощенко П.А. Сверточные нейронные сети с неполной маской весов для задачи распознавания рукописных цифр. Предложена архитектура сверточной нейронной сети с использованием неполной маски весовых коэффициентов в первом слое сети, а также совмещены решения связанные с использованием сверточных слоев с шагом маски более одного и слоев подвыборки. Архитектура спроектирована для решения задачи распознавания рукописных цифр; приведено сравнение работы предлагаемой сети с известными результатами.

Общая постановка проблемы.

Нейронные сети являются одним из мощных и популярных методов для решения задач распознавания образов различной мерности. Однако при применении полносвязной многослойной нейронной сети в ряде задач, сталкиваются с ключевыми недостатками данного вида сети, а именно со значительным ростом размеров сети при увеличении числа входов, а также, нечувствительность к топологии входных данных. Одной из таких задач является задача распознавания рукописных цифр, которая остается по-прежнему актуальной, например, при считывании цифр с квитанций, считывании черновых номеров железнодорожных колес и т.д., когда заказчик желает получить максимальную достоверность. Попыткой уменьшить недостатки традиционной многослойной сети стала разработка сверточной нейронной сети (СНС). В [1] были продемонстрированы возможности сети для работы с изображениями, сигналами и временными рядами. Однако архитектура сети, предложенная в [1] оказалась не совершенной и в [2-5,8] и др. предпринимались различные попытки совершенствования архитектуры сети, для достижения максимального быстродействия и минимальной ошибки распознавания.

Постановка задач исследования.

На сегодняшний день было разработано достаточно большое количество различных архитектур СНС. Ключевыми особенностями являются:

- 1) размер входных данных (в [1] размер входного изображения 28x28 пикселей, в [2-3] – 32x32 пикселей, в [5] – 29x29 пикселей);
- 2) методика передачи данных от слоя к слою (в [5] – полносвязные слои, в [2] используется неполносвязный слой и предложена схема связей между картами слоя);
- 3) метод уменьшения размеров карты нейронов (в [5] предлагается использовать шаг свертки равный 2, за счет чего уменьшить результирующую карту в двое, в [2] уменьшение размеров достигается за счет введения слоев подвыборки);
- 4) тип выходного слоя (в [2,4] выходной слой основан на ядерных функциях, [5] – простой полносвязный слой на основе персептронов).

Разнообразные решения при построении СНС можно трактовать, как попытки человека повлиять на признаки, которые извлекает сеть. В [7] предлагается использование необучаемых сверточных слоев. Данный подход вызван желанием предоставить сети собственноручно выбранные признаки, которые понятны человеку, но при этом лишает ее главного достоинства: гибкости при извлечении признаков путем настройки весов.

В данной работе предлагается новый метод, который использует идею влияния человека на извлекаемые признаки, в контексте задания топологии масок распределенных весовых коэффициентов. Топология маски задается путем указания в маске позиции весов, которые необходимо использовать, а остальные веса маски полагать равными нулю. Целью работы является построение СНС на основе неполных масок для задачи распознавания рукописных цифр. Все исследования проводить на базе данных рукописных цифр MNIST[8].

Решение задачи и результаты исследований.

В основе СНС лежат три архитектурные идеи, которые обеспечивают инвариантность к сдвигу и искажениям. Это – локальные рецептивные поля, распределенные или общие веса и система подвыборки для уменьшения размера карт.

Локальные рецептивные поля представляют собой простейшие обнаружители свойств, основанные на маске весовых коэффициентов, и представляют собой свертку с картой предыдущего слоя. Картой называют группу нейронов, которые объединяют общие веса и пространственные связи. Группа карт одного уровня объединяется в слой. Размеры карты уменьшаются от слоя к слою до тех пор, пока она не вырождается в единственный нейрон. Тем самым карты образуют своеобразную пирамиду. Процесс формирования элемента сверточного слоя показан на рис. 1.

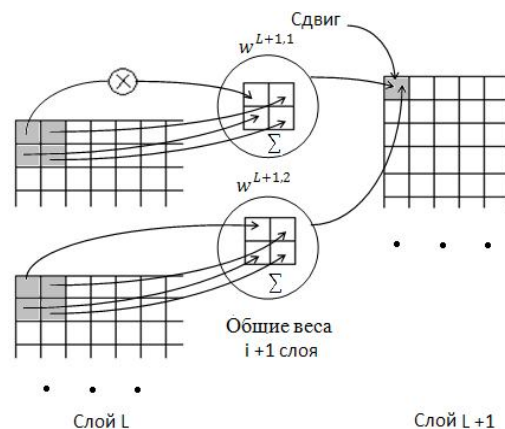


Рисунок 1. Схематическое представление расчета одного элемента сверточного слоя L+1, для маски весов размером 2x2

Аналитически процесс получения элемента сверточного слоя можно записать следующим образом

$$x_{i,j}^{L+1,n} = f \left(\sum_{k=0}^{N^L} \sum_{l=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} x_{l,t}^{L,k} w_{i-l,j-t}^{L+1,k,n} + b^{L+1,n} \right) \quad (1)$$

где $x_{i,j}^{L+1,n}$ - нейрон с координатами (i;j) в n-й карте слоя L+1; $f(\cdot)$ - активационная функция; N^L - число карт в слое L; $w_{i-l,j-t}^{L+1,k,n}$ веса n-й карты слоя L +1, связанные с k-ой картой предыдущего слоя; $b^{L+1,n}$ – сдвиг карты n в слое L +1.

Слой подвыборки имеет такое же количество карт, как и предыдущий слой. Причем карты связаны соотношением 1 к 1. Каждая карта слоя подвыборки имеет один весовой коэффициент и один сдвиговый. Элемент слоя подвыборки формируется путем суммирования значений выходов нейронов в некоторой окрестности (обычно 2x2) предыдущего слоя с умножением на весовой коэффициент и добавлением сдвига. Окрестности для любых 2-х отдельно взятых нейронов слоя подвыборки не пересекаются, а в сумме покрывают всю карту.

Важним елементом архітектури являється розмір маски. Для даних цілей испытывали маски 3x3, 5x5 и 7x7 (например в[5]). В качестве наилучшей маски, в большинстве публикаций была выбрана маска 5x5, поскольку маска 3x3 не дает желаемого результата, а маска 7x7 слишком велика для входного изображения 28x28 базы MNIST.

Применение распределенных масок или рецептивных полей в первую очередь связано с попыткой извлекать элементарные пространственные связи. Однако влияние на выбор этих связей можно оказать только процессом обучения и инициализации весов. Для усиления это влияния в работе предлагается использовать неполные маски различной структуры.

Далее в качестве исходного размера маски была выбрана маска 5x5. Так как в масках предполагается не использовать некоторые веса, то, для предотвращения возможной потери данных, необходимо использовать маски различной структуры. Структуры предлагаемых масок показаны на рис 2.

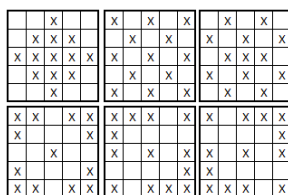


Рисунок 2. Маски, которые использованы для формирования пространственных связей в рецептивном поле (X- наличие весового коэффициента)

Как видно из рис.2, для всех масок было установлено по 13 элементов. Это обусловлено тем, что выбор меньшего числа элементов маски (например, 9) не предоставляет возможность сконфигурировать маски желаемой структуры. А при выборе большего числа элементов (например, 17) наблюдается их избыточность. Данные маски предлагается использовать в первом сверточном слое СНС. Для задачи распознавания рукописных цифр была разработана архитектура СНС, которая показана на рис 3.

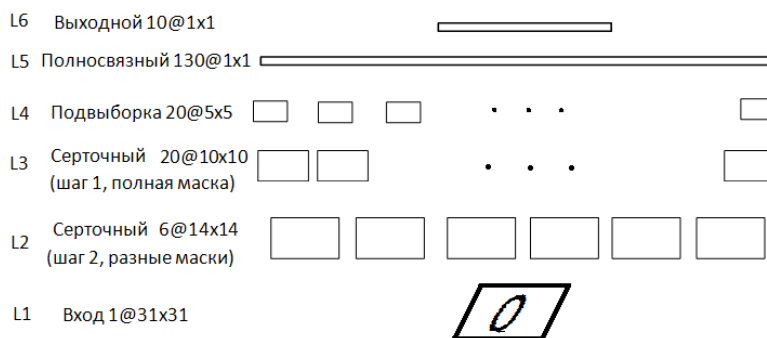


Рисунок 3. Архитектура прилагаемой СНС (формат записи структуры слоя a@b означает, что a- количество карт, b - размеры карты)

Рассмотрим предлагаемую архитектуру подробнее. Входной слой L1 имеет размер 31x31. Данный размер определяется двумя параметрами: размером изображения символа 28x28 и структурой слоев следующих уровней. Так как изображения в базе MNIST имеют размер 28x28, то для приведения к размеру 31x31 их необходимо дополнить фоновым значением. С предварительным центрированием изображений символов по каждой координате отдельно. Такое преобразование необходимо для того, чтобы при обучении можно было применять сдвиговые преобразования в равном диапазоне по каждому направлению.

Сверточный слой L2 состоит из 6 карт размером 14x14. Маска весов каждой карты соответствует одной из масок приведенных на рис 2. Таким образом, мы имеем 1176

нейронов (по 196 нейронов в карте) и $13 \cdot 6 + 6 = 84$ настраиваемых параметра (6 сдвигов и 78 весов) в слое. Шаг проведения свертки равен двум (аналогично [5]), фактически после проведения свертки, каждая вторая строка и каждый второй столбец выбрасываются.

Следующий слой L3 также сверточный с шагом 1, но не полносвязный. Он состоит из 20 карт размером 10×10 . Маска свертки – полная. Каждая карта слоя L3 соединена с тремя картами предыдущего слоя. Схема соединений приведена в таблице 1. В слое 2000 нейронов, и $20 \cdot 25 \cdot 3 + 20 = 1520$ настраиваемых параметров.

Таблица 1. Соединения карт слоя L3 с картами слоя L2. Номера карт слоя L3 не имеют значения.

L3 вход	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	x	x	x	x	x	x	x	x	x											x
2	x	x	x	x						x	x	x	x	x					x	
3	x				x	x	x			x	x	x			x	x				x
4		x			x			x	x	x			x	x	x	x	x			
5			x			x		x			x		x		x		x	x	x	x
6				x			x		x			x		x		x	x	x	x	x

Слой подвыборки L4, в силу своих особенностей имеет, как и L3, 20 карт размером 5×5 . В слое 500 нейронов и 40 обучаемых параметров.

Полносвязный слой L5 имеет 130 вырожденных карт (когда карта состоит из одного нейрона). В слое 130 нейронов и $130 \cdot 25 \cdot 20 + 130 = 65130$ обучаемых параметров. Выходной слой также является полносвязным и содержит 10 нейронов и $10 \cdot 130 + 10 = 1310$ обучаемых параметров.

Итого в предложенной сети 3816 нейронов и 68084 обучаемых параметра.

Инициализация весов и сдвигов проводилась случайными значениями со стандартным отклонением 0.08. Входные и выходные параметры преобразовывались к диапазону $[-1; 1]$. В качестве активационной функции была выбрана масштабированная функция гиперболического тангенса:

$$f(a) = A \tanh(Sa) \quad (2)$$

где $f(a)$ – искомое значение; a – взвешенная сумма сигналов; A , S – параметры функции. В [3] для данных параметров были предложены следующие значения: $A = 1.759$ и $S = 2/3$. Параметр A выбран таким образом, чтобы диапазон выходных значений попадал на линейную часть активационной функции.

При обучении, на каждой эпохе обучающая выборка подвергалась масштабированию, сдвигу и повороту со случайными значениями для каждого экземпляра. Процесс обучения проводился по алгоритму обратного распространения ошибки с учетом вторых производных. Детально алгоритм описан в [2,3], а пример реализации можно найти в [6].

Особенность предложенной реализации алгоритма СНС (в отличие от [6] и аналогичных) заключается в отказе от явного указания точек соединения нейронов карты одного слоя с нейронами карты другого слоя. Вместо этого применяется дополнительная карта смещений, в которой по одной оси указываются индекс веса в маске, по другой – номер выходного нейрона, а на пересечении номер нейрона в карте предыдущего слоя, с которым необходимо проводить свертку. Это возможно благодаря тому, что каждая карта представляется в памяти виде линейного массива. Применение карты смещений дало возможность существенно сократить расходы на использование памяти без потерь производительности, для случая, когда сеть работает в прямом режиме (распознавания), и дало небольшой выигрыш в вычислительной сложности при обучении, что позволило

сократить время одной эпохи обучения на 5%, по сравнению с аналогичной реализацией, но с использованием явных связей.

Сравнение результатов работы предложенной сети для задачи распознавания тестирующей выборки базы MNIST приведено в таблице 2.

Таблица 2. Сравнение результатов работы

Алгоритм	Тип искажения обучающей выборки	Количество обучаемых параметров (только у СНС)	Ошибка	Ссылка на работу, из которой взят результат
2 layer MLP(MSE)	аффинное	-	1.6%	[2]
Tangent dist.	аффинное + утолщение	-	1.1%	[2]
Lenet5 (MSE)	аффинное	60550	0.8%	[2]
Boost. Lenet4 MSE	аффинное	-	0.7%	[2]
2 layer MLP (CE)	эластическое	-	0.7%	[5]
Simple conv (CE)	аффинное	133816	0.6%	[5]
Simple conv (CE)	эластическое	133816	0.4%	[5]
СНС с неполной маской	аффинное	68084	0.44≈0.4%	данная работа

Выводы.

Предложенная СНС, которая отличается от известных тем, что использует маски с неполными весами, а также, использованием в одной сети как слоев с шагом маски более одного, так и слоев подвыборки. Такое построение дало возможность показать результат сравнимый с лучшим на сегодняшний день показателем, и при этом достигать результат с использованием меньшего объема ресурсов, в том числе и процессорных, что приводит к более высокому быстродействию. Кроме того, результат удалось достичь без применения эластических искажений.

Список литературы

1. ЛиКун Я., Бэнжио Я. Convolutional Networks for Images, Speech, and Time-Series in Arbib, M. A. (Eds), The Handbook of Brain Theory and Neural Networks, MIT Press, 1995.
2. ЛиКун Я., Бэнжио Я., Хаффер П. Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.
3. ЛиКун Я., Ботто Л., Опп Г., Мюллер К. "Efficient BackProp". Neural Networks: Tricks of the trade, 1998.
4. Миrowski П., ЛиКун Я., Мадхаван Д., Кузнецкий Р. Comparing SVM and Convolutional Networks for Epileptic Seizure Prediction from Intracranial EEG. Proc. Machine Learning and Signal Processing (MLSP'08), IEEE, 2008.
5. Симард П., Стеинкраус Д., Платт Дж., Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. International Conference on Document Analysis and Recognition (ICDAR), IEEE Computer Society, Los Alamitos, pp. 958-962, 2003.
6. Онеил М. Neural Network for Recognition of Handwritten Digits. <http://www.codeproject.com/KB/library/NeuralNetRecognition.aspx>
7. Царегородцев В.Г. Сверточные нейронные сети с необучаемыми сверточными слоями. Международная конференция «Вычислительные и информационные технологии в науке, технике и образовании», 2008г
8. ЛиКун Я., The MNIST database of handwritten digits. <http://yann.lecun.com/exdb/mnist>.