

УДК 004.[932.75'1+89]

## ПРИМЕНЕНИЕ КОНФИГУРАЦИОННОГО КОДИРОВАНИЯ ВЫХОДНОГО СИГНАЛА В НЕЙРОННОЙ СЕТИ СВЁРТКИ ДЛЯ РАСПОЗНАВАНИЯ РУКОПИСНЫХ СИМВОЛОВ



**Я.А. Бурый**

*Ассистент кафедры электронных вычислительных машин БГУИР, аспирант кафедры электронных вычислительных машин БГУИР*



**Д.И. Самаль**

*Доцент кафедры электронных вычислительных машин факультета компьютерных систем и сетей БГУИР, кандидат технических наук, доцент*

*Белорусский государственный университет информатики и радиоэлектроники  
Email: edidici@tut.by, samal@bsuir.by*

**Аннотация.** Статья посвящена вопросу применения конфигурационного кодирования выходного сигнала нейронной сети свёртки при распознавании большого количества классов с целью уменьшения её сложности. Несмотря на то, что свёрточные нейронные сети неплохо справляются с задачей распознавания образов, существуют типы задач, при которых их применение ограничено чересчур большим количеством классов распознаваемых образов, что в итоге крайне отрицательно сказывается как на времени обучения, так и непосредственно распознавания. Известен конфигурационный способ кодирования выходного сигнала нейронных сетей, однако, область его применения, как правило, ограничена распознаванием различных сопутствующих признаков образов, таких как местоположение образа на входном изображении, вытянутость образа и т.п. В рассматриваемой работе конфигурационное кодирование применяется непосредственно для представления кода распознаваемого символа на выходе нейронной сети, что позволяет существенно снизить её общую сложность в случае значительного количества классов распознаваемых символов, начиная от нескольких сотен и более.

**Ключевые слова:** нейронные сети, нейронная сеть свёртки, позиционное кодирование, конфигурационное кодирование, гибридное конфигурационно-позиционное кодирование, распознавание символов, большие объёмы данных, MNIST.

**Введение.** Свёрточные нейронные сети в последние годы хорошо зарекомендовали себя в качестве классификаторов, используемых для распознавания изображений в силу их хорошей обобщающей способности, а также некоторой инвариантности к афинным преобразованиям и искажающим деформациям входного сигнала.

Применительно к системам распознавания возможны три способа представления выходного сигнала:

1. *Позиционное кодирование*, при котором код принимает форму двоичного числа, все разряды которого, за исключением одного, состоят из нулей. Положение ненулевого разряда указывает на класс идентифицируемого стимула [1]. При этом общее количество выходных нейронов соответствует количеству распознаваемых классов образов  $M = N$ , где  $N$  – количество распознаваемых классов образов, а  $M$  – количество выходных нейронов сети.

При позиционном кодировании вероятность правильной идентификации входного сигнала наибольшая, т.к. в этом случае примеры из обучающей выборки, которые сопоставляются с конкретным выходом нейронной сети, располагаются наиболее компактно в пространстве признаков вследствие близости образов, принадлежащих одному и тому же классу, а соответствующие им многообразия содержат минимальное количество краёв и разрывов. Вследствие этого данный способ кодирования применяется при распознавании непосредственно классов самих образов, а не каких-либо их вспомогательных признаков [1].

2. *Конфигурационное кодирование*, при котором каждой реакции сети приписывается некоторое двоичное число [2]. При таком способе кодирования количество выходных нейронов значительно меньше и составляет  $M = \log_2 N$ . При этом каждому конкретному выходу нейронной сети сопоставляется примерно половина распознаваемых классов образов и, следовательно, столько же самих изображений, содержащихся в обучающей выборке в случае, если последняя содержит равное количество примеров для каждого класса. Это следует из того, что вероятность принять значение 1 или 0 для каждого бита случайно взятого двоичного числа любой длины равняется 0,5. А значит каждый выход системы распознавания реагирует на множество, с высокой вероятностью содержащее несколько краёв, а также разрывы, и в общем случае представляющее собой гауссову смесь, для которой не всегда возможно однозначно определить линию огибающей, что должно отрицательно сказываться на качестве распознавания.

Данный способ кодирования обычно применяется для определения какого-либо независимого признака или свойства стимула, такого, как местоположение объекта на входном изображении, его размер, вытянутость в горизонтальном или вертикальном направлении и т.п. [2].

3. *Гибридное конфигурационно-позиционное кодирование*. Здесь применяется смешанный подход, при котором нейронная сеть обучается формировать заданные выходные последовательности, определяющие конкретные классы распознаваемых образов. Затем на основе анализа полученной последовательности делается вывод о принадлежности поданного на вход изображения к тому или иному классу.

Количество выходных нейронов для данного типа кодирования может варьироваться, и определяется выбранной длиной  $K$  выходной последовательности, генерируемой нейронной сетью, с тем учётом, чтобы количество всех возможных значений таких последовательностей было не меньше, чем количество распознаваемых классов. Как и в случае применения конфигурационного кодирования, здесь каждый выход нейронной сети в общем случае также реагирует на многообразие в виде гауссовой смеси, однако, её конкретная форма и, в частности, количество краёв и разрывов зависят от длины  $K$ , которая может варьироваться в диапазоне  $\log_2 N \leq K \leq N$ . При этом достигаемая точность системы распознавания предположительно также будет соответственно изменяться.

Для повышения скорости распознавания обработке может подвергаться не всё изображение целиком, а некоторое количество его точек, как, например, было сделано в работе [3] для алгоритма восстановления изображений.

Данный тип кодирования выходного сигнала применяется в случаях, когда число классов значительно, например, несколько сотен и более, и позволяет добиться определённого баланса между временем работы и качеством распознавания нейронной сети.

**Постановка задачи.** Существует немало задач, при которых размер множества распознаваемых классов образов существенно ограничивает применение позиционного кодирования выходного сигнала, достигая в некоторых случаях десятков или даже сотен тысяч и более. Вот примеры лишь некоторых из них:

1. Китайские, японские и т.п. иероглифы;
2. Множества видов животных, растений и т.п.;

3. Множества слов (морфем) любого языка в случае распознавания целыми словами (морфемами), а не посимвольно;
4. Изображения лиц различных людей;
5. Изображения отпечатков пальцев различных людей;
6. Изображения участков звёздного неба.

Для определения предельных возможностей минимизации выходного сигнала, а также влияния такой минимизации на общую сложность системы распознавания и её результирующую точность, была выбрана архитектура нейронной сети свёртки, описанная в таблице 1. Она достаточно проста, но при этом показывает отличные результаты при распознавании небольших двоично закодированных изображений.

Таблица 1

Описание параметров нейронной сети свёртки

Но- мер слоя	Тип слоя	Размер карты при- знаков	Количество карт призна- ков	Размер ре- цептивного поля нейрона	Общее ко- личество нейронов	Общее ко- личество связей
0	Входной	28 x 28	1	0	784	0
1	Свёрточный	24 x 24	6	5 x 5 x 1	3 456	86 400
2	Подвыбороч- ный	12 x 12	6	2 x 2	864	3 456
3	Свёрточный	8 x 8	16	5 x 5 x 6	1 024	153 600
4	Подвыбороч- ный	4 x 4	16	2 x 2	256	1 024
5	Полносвязный выходной	1 x 1	10	4 x 4 x 16	10	2 560
Итого:					6 394	247 040

При тестировании описанной нейронной сети для распознавания рукописных символов из базы MNIST[4] с использованием позиционного кодирования выходного сигнала была достигнута точность 98,21% на обучающей выборке и 97,13% на тестовой выборке. Среднеквадратичная ошибка выходов сети при этом составила 0,1108 на обучающей выборке и 0,1247 на тестовой выборке.

Согласно данным из таблицы 1, общее количество связей выбранной нейронной сети составляет 247 040. При этом  $S_{1-4} = 86\,400 + 3\,456 + 153\,600 + 1024 = 244\,480$  связей приходится на внутренние слои от 1-го по 4-й и только  $S_5 = 2\,560$  связей – на выходной 5-й слой. Количество связей выходного слоя здесь зависит не только от количества выходов всей нейронной сети, но также от количества выходов предпоследнего слоя и равно  $S_5 = N_4 \cdot N_5 = 256 \cdot 10 = 2560$ , где  $N_4$  и  $N_5$  – количество выходов, соответственно 4-го и 5-го слоя.

Отсюда видно, что общее количество связей нейронной сети растёт по линейному закону и для произвольного количества выходов  $N_5$  составит  $S_5 = S_{1-4} + N_4 \cdot N_5 = (86\,400 + 3\,456 + 153\,600 + 1024) + 256 \cdot N_5 = 244\,480 + 256 \cdot N_5$ .

В таблице 2 представлены результирующие значения количества связей данной нейронной сети при различных значениях количества выходов, а также относительное увеличение количества связей.

Как видно из таблицы 2, при существенном увеличении количества распознаваемых классов общая сложность нейронной сети может значительно увеличиваться – в десятки, сотни и тысячи раз, что, теоретически, будет иметь существенное влияние на скорость обучения и распознавания.

В таких случаях целесообразно попытаться применить конфигурационное кодирование не только к распознаванию независимых признаков, но и непосредственно – самих

образов, учитывая, что в таком случае количество связей нейронной сети будет расти по логарифмическому закону, и для описанной архитектуры, при количестве распознаваемых классов образов  $N = 1\,000\,000$ , составит  $S_{1-5} = S_{1-4} + 256 \cdot \log_2 N = 244\,480 + 256 \cdot 20 = 249\,600$ , что, по сравнению с исходной архитектурой, увеличивает количество связей всего в 1.01 раза.

Таблица 2

Сложность нейронной сети свёртки для различного количества классифицируемых образов

Количество распознаваемых классов образов	Результирующее количество связей	Относительное увеличение количества связей, раз
10	$244\,480 + 256 \cdot 10 = 247\,040$	1
100	$244\,480 + 256 \cdot 100 = 270\,080$	$\approx 1,1$
1 000	$244\,480 + 256 \cdot 1\,000 = 500\,480$	$\approx 2$
10 000	$244\,480 + 256 \cdot 10\,000 = 2\,804\,480$	$\approx 11,4$
100 000	$244\,480 + 256 \cdot 100\,000 = 25\,844\,480$	$\approx 104,6$
1 000 000	$244\,480 + 256 \cdot 1\,000\,000 = 256\,244\,480$	$\approx 1037,3$

Целью настоящего исследования является оценка падения точности нейронной сети и увеличение среднеквадратической ошибки её выхода при конфигурационном кодировании выходного сигнала.

**Результаты экспериментов.** Для проведения эксперимента были выбраны два различных способа представления выходного сигнала системы распознавания: конфигурационно-пороговое и конфигурационно-конкурентное кодирование. В первом случае для количества классифицируемых образов  $N = 10$  использовалось 4 выходных нейрона, результирующий сигнал которых для получения двоичного кода целого числа корректировался по следующей формуле:  $y'_i = \begin{cases} 1, & \text{если } y_i > 0 \\ 0, & \text{если } y_i \leq 0 \end{cases}$ , где  $y'_i$  – значение  $i$ -го выходного нейрона после коррекции, а  $y_i$  – значение  $i$ -го выходного нейрона до коррекции. Во втором случае использовалось 8 выходных нейронов, при этом каждая пара нейронов кодировала 1 бит выходного сигнала по формуле:  $y'_i = \begin{cases} 1, & \text{если } y_{i1} = \max(y_{i1}, y_{i2}) \\ 0, & \text{если } y_{i2} = \max(y_{i1}, y_{i2}) \end{cases}$ , где  $y_{i1}$  и  $y_{i2}$  – пара выходных нейронов, кодирующих  $i$ -й бит выходного сигнала.

Для достижения цели данного исследования была выбрана база изображений рукописных символов MNIST, т.к. она является широко используемым стандартным средством проверки систем распознавания. В ней содержится 60 000 обучающих и 10 000 тестовых изображений арабских цифр от 0 до 9.

Ввиду того, что использование конфигурационно-порогового и конфигурационно-конкурентного кодирования для 10 указанных классов не влияет существенно на итоговую сложность нейронной сети ( $244\,480 + 256 \cdot 4 = 245\,504$  и  $244\,480 + 256 \cdot 8 = 246\,528$  связей, соответственно, против 247 040 связей при позиционном кодировании), влияние применения различных способов кодирования на время обучения и распознавания не приводится, однако, эти параметры, очевидно, находятся в прямой зависимости от итогового количества связей нейронной сети и могут быть примерно оценены на основании данных из таблицы 2.

Достигнутые значения среднеквадратической ошибки распознавания, абсолютной ошибки распознавания и абсолютной точности распознавания представлены в таблице 3.

Таблица 3.

Результаты обучения нейронной сети на базе рукописных символов MNIST

Кодирование выходного сигнала	Среднеквадратическая ошибка сети		Абсолютная ошибка распознавания		Абсолютная точность распознавания	
	На обучающей выборке	На тестовой выборке	На обучающей выборке	На тестовой выборке	На обучающей выборке	На тестовой выборке
Позиционное	0,1108	0,1247	0,0179	0,0287	98,21%	97,13%
Конфигурационно-пороговое	0,1635	0,1796	0,0827	0,1039	91,73%	89,61%
Конфигурационно-конкурендное	0,1462	0,1574	0,0579	0,0698	94,23%	93,02%

Результаты работы алгоритма обучения нейронной сети в динамике для различных способов кодирования выходного сигнала представлены на рисунке 1.

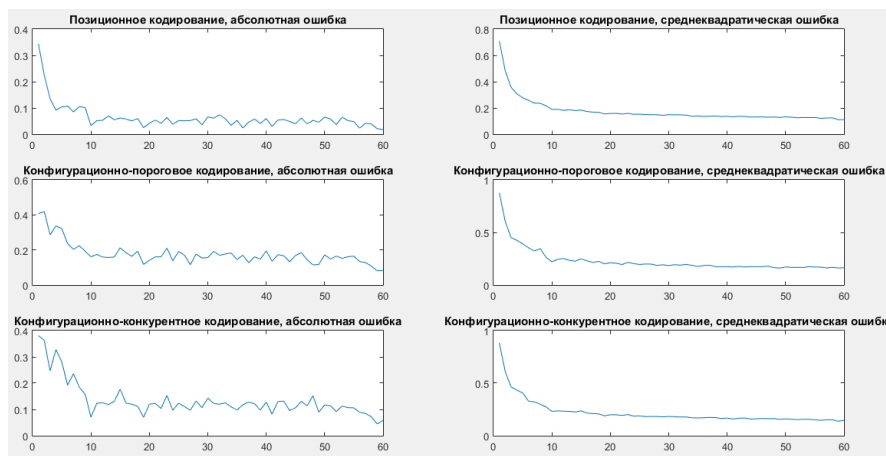


Рисунок 1. Динамические результаты работы алгоритма обучения нейронной сети для 60 эпох обучения

**Заключение.** В ходе проведённого эксперимента было определено, что падение точности распознавания по сравнению с позиционным кодированием выходного сигнала в случае применения конфигурационно-порогового кодирования составило 6,48% на обучающей и 7,52% на тестовой выборке, а в случае конфигурационно-конкурендного кодирования – 3,98% на обучающей выборке и 4,11% на тестовой выборке.

Это указывает на то, что второй способ кодирования выходного сигнала приводит к более точным результатам и лучше подходит для применения в задачах, связанных с распознаванием большого количества классов образов без существенного уменьшения скорости обучения и распознавания. При необходимости повышения точности распознавания может быть применена некоторая гибридизация кодирования выходного сигнала путём введения дополнительных контролируемых выходов нейронной сети, наподобие того, как это используется в схемотехнических решениях. При этом размерность выходного сигнала будет увеличена на константное количество бит, однако при этом будет сохранён логарифмический закон роста сложности системы распознавания.

### *Литература*

- [1] Позиционное кодирование - [Электронный ресурс]. - Режим доступа: [https://ru.wikipedia.org/wiki/Позиционное\\_кодирование](https://ru.wikipedia.org/wiki/Позиционное_кодирование)
- [2] Конфигурационное кодирование - [Электронный ресурс]. - Режим доступа: [https://ru.wikipedia.org/wiki/Конфигурационное\\_кодирование](https://ru.wikipedia.org/wiki/Конфигурационное_кодирование)
- [3] Яковлев С. С. Система распознавания движущихся объектов на базе искусственных нейронных сетей // ИТК НАНБ. — Минск, 2004. — С. 230—234.
- [4] База изображений THE MNIST DATABASE of handwritten digits - [Электронный ресурс]. - Режим доступа: [http:// http://yann.lecun.com/exdb/mnist/](http://yann.lecun.com/exdb/mnist/). – Дата доступа: 01.04.2018.
- [5] Прохоров В.Г. Использование сверточных нейронных сетей для распознавания рукописных символов // УкрПрог, Киев, Украина. 2008.
- [6] Хайкин С. Нейронные сети. Полный курс. / М., СПб., Киев: Вильямс, 2006. – 1104с.

## **APPLYING OF THE CONFIGURATIVE CODING OF THE OUTPUT SIGNAL IN THE NEURAL NETWORK FOR RECOGNITION OF HAND-HANDED SYMBOLS**

***Y.A. BURYI***

*PhD Student, Assistant of the Department of  
Electronic Computing Machines  
of the Faculty of Computer Systems and Net-  
works of BSUIR*

***D.I. SAMAL***

*PhD (Candidate of Technical Sciences),  
Associate Professor of the Department of Elec-  
tronic Computing Machine of the Faculty of  
Computer Systems and Networks  
of BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus  
E-mail: edidici@tut.by , samal@bsuir.by*

**Abstract.** The article is devoted to applying of configuration coding of the output signal of the convolution neural network in the recognition of a large number of classes in order to reduce its complexity.

**Keywords:** neural networks, neural network convolutions, positional coding, configuration coding, hybrid configurational-positional coding, character recognition, big data, MNIST.