

УДК 621.391

В.Г. ИВАНОВ<sup>1</sup>, С.А. ЧЕРВЕНКО<sup>2</sup><sup>1</sup>Национальная юридическая академия Украины им. Я. Мудрого,<sup>2</sup>Национальный аэрокосмический университет им. Н. Жуковского «ХАИ», Украина

## КОДИРОВАНИЕ ДАННЫХ ПО СИСТЕМЕ ХААРА В НЕЙРОСЕТЕВОЙ СРЕДЕ

Приводятся результаты исследования возможностей применения искусственной нейросреды для кодирования цифровых данных коэффициентами ортогонального ряда по системе функций Хаара. Показано, что учитывая линейный характер и специфику дискретного преобразования Хаара, отражающуюся в функции распределения элементов выходного вектора (большое количество значений близких к 0), лучшие результаты могут быть получены при использовании двухслойной нейронной MLP-структуры с сигмоидальными передаточными функциями для скрытого слоя и линейными для второго слоя

нейросетевая среда, преобразование Хаара, сжатие данных, искусственная нейронная сеть

### Введение

Эффективное кодирование данных различной физической природы являлось и является основной задачей всех систем получения, передачи и хранения информации. Достаточно эффективным способом кодирования зарекомендовал себя метод замены исходных отсчетов вектора данных коэффициентами того или иного ортогонального преобразования, выбираемого по различным критериям, например по скорости получения коэффициентов. Такой подход используется сегодня в системах сжатия и кодирования речевых сигналов и изображений, системах классификации и распознавания, системах технического зрения и шифрования [1]. Оптимальным, в смысле скорости получения ортогональных коэффициентов, является система нестационарных кусочно-постоянных функций базиса Хаара, который демонстрирует фундаментальную связь математической теории и практики обработки сигналов, проявляющейся в широком использовании вейвлет-преобразований, базовой идеологией которых являются правила построения и вычислений функций в системе Хаара. Поэтому весьма актуальной является задача исследования возможностей применения искусственной нейросреды (ИНС) для кодирования цифровых данных коэффициентами ортогонального ряда по системе функций Хаара.

### 1. Постановка задачи

Общая постановка решаемой задачи выглядит следующим образом: необходимо построить отображение, такое, чтобы на каждый возможный входной сигнал  $X$  формировался правильный выходной сигнал  $Y$ . Отображение задается конечным набором пар (<вход>, <известный выход>). Число таких пар (обучающих примеров) существенно меньше общего числа возможных сочетаний значений входных и выходных сигналов. Совокупность всех обучающих примеров – это обучающая выборка. В нашей задаче  $X$  – подмножество цифровых RGB-отсчетов изображения, являющееся входным вектором данных преобразователя Хаара, а  $Y$  – формируемые этим преобразователем выходные данные соответствующей размерности. Отображение

$$g : X \rightarrow Y$$

является многомерным, т.е.  $X$  и  $Y$  – вектора с размерностью больше единицы.

### 2. Решение задачи

Учитывая линейный характер и специфику дискретного преобразования Хаара [2], проявляющуюся в функции распределения элементов выходного вектора (большое количество значений, близких к 0), лучшие результаты могут быть получены при использовании двухслойной ИНС MLP-структуры с

сигмоидальными передаточными функциями для скрытого слоя и линейными для второго слоя.

Известно, что для обучения таких ИНС, используются алгоритмы обучения с учителем [3]. Задача обучения определяется совокупностью 5 элементов:

$$\langle X, Y, r, G, E \rangle,$$

где  $X$  и  $Y$  – вход и выход соответственно;  $r$  – функция – определяет желаемый результат обучения; в задаче обучения по примерам функция  $r$  задается парами входных-выходных данных:  $(X_1, Y_1)$ ;  $(X_2, Y_2)$ ; ... ;  $(X_M, Y_M)$ , для которых  $Y_m = r(X_m)$  ( $m = 1, 2, \dots, M$ ); архитектура связей нейронной сети считается заданной до начала обучения, она определяет множество функций  $G$ ;  $E$  – функция ошибки, показывающая для каждой функции степень близости к  $r$ ; обучение состоит в поиске (синтезе) функции  $g$ , оптимальной по  $E$  [3, 4].

Для обучения ИНС [5, 6] мы использовали оптимальный алгоритм Левенберга-Марквардта, способный обеспечить наиболее высокую скорость сходимости.

Рассмотрим нейросетевую реализацию аппроксимации функции дискретного преобразования Хаара отсчетов яркости изображения «Zelda». Пусть имеется множество пар векторов, где каждый входной вектор объединяет 8 отсчетов исходного изображения, а соответствующий ему выходной – 8 вещественных чисел – результатов дискретного преобразования Хаара входного вектора. Размерность входного и выходного векторов определяет размерность входных и выходных данных и количество нейронов в выходном слое синтезируемой структуры ИНС.

Диапазоны входных данных (Input ranges) необходимо определить исходя из подготовленной выборки. В соответствии с выбранной структурой ИНС, первая, созданная нами, MLP-структура имеет 2 слоя, с линейными функциями активации (Purelin) в выходном слое и гиперболическими тангенсами (TanSig) сигмоидального вида в скрытом.

Строго определенной процедуры для выбора количества нейронов в скрытом слое ИНС нет. Чем больше количество нейронов и слоев, тем шире возможности сети, тем медленнее она обучается и работает и тем более нелинейной может быть зависимость вход-выход.

Если ИНС в скрытом слое содержит слишком мало нейронов, то:

- сеть не обучится и ошибка при работе сети останется слишком большой;
- на выходе сети не будут передаваться резкие колебания аппроксимируемой функции  $y(x)$ .

Превышение требуемого количества нейронов тоже мешает работе сети. Если нейронов слишком много возникают следующие проблемы:

- быстродействие будет низким и потребуется много ресурсов для обучения;
- сеть переобучится и выходной вектор будет передавать незначительные и несущественные детали в изучаемой зависимости  $y(x)$ , например, ошибочные данные;
- зависимость выхода от входа окажется резко нелинейной: выходной вектор будет существенно и непредсказуемо меняться при малом изменении входного вектора  $x$ ;
- сеть будет неспособна к обобщению: в области, где нет или мало известных точек функции  $y(x)$  выходной вектор будет случаен и непредсказуем, не будет адекватен решаемой задаче.

Следовательно, оптимальное значение должно быть найдено опытным путем исходя из требований к точности аппроксимации заданной функции.

Было сформировано множество из 5000 пар векторов, где входные вектора объединяют по 8 отсчетов яркости изображения «Zelda», а соответствующие им выходные – результаты дискретного преобразования Хаара. Подмножество для создания обучающей выборки из 2500 пар векторов формировалось посредством случайного выбора из исходного множества с целью внесения наибольшего возмож-

ного разнообразия в данные, на которых будет обучаться модель ИНС.

Теорией количество пар в обучающем множестве не регламентируется, но если элементов слишком много или мало, ИНС не обучится и не решит поставленную задачу [7]. Количество элементов в обучающем множестве должно быть достаточным для обучения сети, чтобы под управлением алгоритма сформировать набор параметров сети, дающий нужное отображение  $g: X \rightarrow Y$ .

Данные, подаваемые на вход сети и снимаемые с выхода, должны быть правильно подготовлены. Один из распространенных способов, который и был применен, – масштабирование:

$$x = (x' - m)c, \quad (1)$$

где  $x'$  – исходный вектор,  $x$  – масштабированный, вектор  $m$  – некоторое усредненное значение совокупности входных данных,  $c$  – масштабный коэффициент.

Масштабирование желательно, чтобы привести данные в допустимый диапазон [8]. Если этого не сделать, то может возникнуть несколько проблем:

1) нейроны первого слоя или окажутся в постоянном насыщении ( $|m|$  велик, дисперсия входных данных мала) или будут все время заторможены ( $|m|$  мал, дисперсия мала);

2) весовые коэффициенты примут очень большие или очень малые значения при обучении (в зависимости от дисперсии), и, как следствие, растянется процесс обучения и снизится точность.

В данной реализации масштабирование выполнено по формуле

$$x = \frac{(x' - \min(x'))}{\max(x') - \min(x')}, \quad (2)$$

приводящей значения всех скаляров в интервал  $[0; 1]$ .

Выходы сети масштабируются также. Так как мы сами выбираем смысл выходного вектора при создании сети, то надо подготовить данные так, чтобы

диапазон изменения выходных сигналов лежал на рабочем участке функции активации.

Тестовая выборка для симуляции процесса функционирования ИНС была сформирована при помощи масштабирования по тому же правилу (2) и составлена из всех 5000 пар векторов исходного множества с сохранением их последовательности.

Обычно скорость обучения зависит от времени обучения  $\varepsilon(t)$ , функции, монотонно убывающей с ростом времени. Для сходимости алгоритма необходимо:

$$\varepsilon(t) \xrightarrow{t \rightarrow 0} 0; \quad \int_0^{\infty} \varepsilon(t) dt = +\infty. \quad (3)$$

Результаты обучения показали, что пакетному алгоритму Левенберга-Марквардта достаточно 15 итераций, чтобы уменьшить обобщенный показатель среднеквадратичной ошибки на 3 порядка для выбранной нейроструктуры, после чего изменение стабилизируется, что говорит о достижении оптимума (рис. 1).

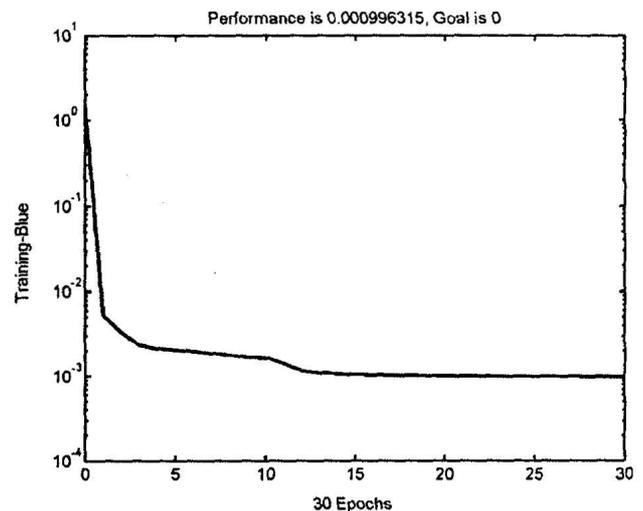


Рис. 1. Динамика ошибки при обучении (метод Левенберга-Марквардта)

Для получения более точной аппроксимирующей модели была также сгенерирована сеть с 8 нейронами в скрытом слое и линейными выходными функциями активации. Динамика процесса обучения этой модели ИНС отображена на графике (рис. 2).

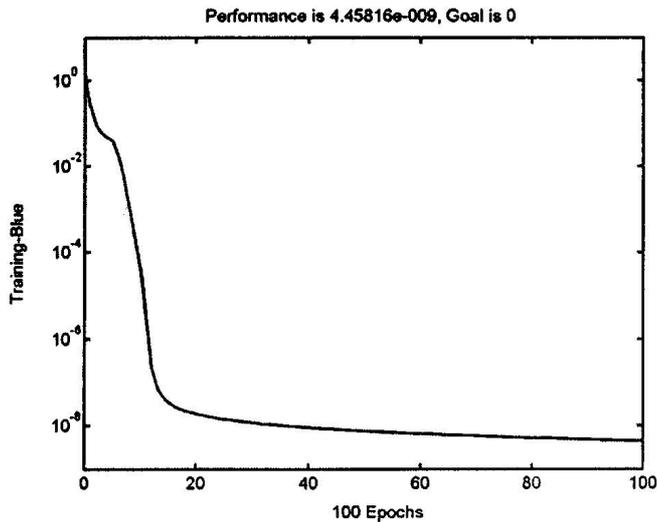


Рис. 2. Динаміка навчання моделі з 8 нейронами в прихованому шарі

Для навчання данної моделі на підготовленій вибірці потрібно було вже около 100 ітерацій алгоритма, різко возросло і час навчання на одному кроці, так як метод Левенберга-Марквардта використовує в розрахунках матрицю, розмірність якої визначається кількістю параметрів мережі. Але зменшити узагальнений показник середньоквадратичної помилки вдалося на 5 порядків більше, в порівнянні з структурою з 4 нейронами прихованого шару.

### 3. Результати моделювання

Порівняльні результати тестування моделей з 4 і 8 нейронами в прихованих шарах відображені в табл. 1.

Розрахунок даних, розміщених в табл. 1, виконувався виходячи з значень помилки апроксимації після ренормалізації отриманих даних, а не помилок виходів ІНС, що дозволяє легше оцінювати «дійствительну» точність моделі. Крім усередненого і максимального значення модуля помилки по кожному елементу вихідного вектора в табл. 1 також вказано корінь із середньоквадратичної помилки – показник, найбільш часто застосовуваний для відображення ефективності функціонування ІНС.

Таблиця 1

Результати тестування двох моделей ІНС

Out №	RMSE		Mean abs(E)		Max abs(E)	
	Model 8-4-8	Model 8-8-8	Model 8-4-8	model 8-8-8	model 8-4-8	model 8-8-8
1	0,2115	0,0166	0,1329	0,0102	02,0876	0,3047
2	1,8255	0,0067	1,1964	0,0036	25,6368	0,1181
3	1,1901	0,0039	0,8019	0,0020	10,5071	0,0807
4	1,4394	0,0060	0,9574	0,0022	16,8550	0,2384
5	1,8122	0,0029	1,1526	0,0013	14,5978	0,0880
6	1,3919	0,0023	0,9521	0,0010	11,1751	0,0581
7	1,2438	0,0029	0,8668	0,0011	15,7700	0,1163
8	0,9693	0,0023	0,6372	0,0012	12,7417	0,0366
Mean	<b>1,2605</b>	<b>0,0055</b>	<b>0,8371</b>	<b>0,0028</b>	<b>13,6714</b>	<b>0,1301</b>

Усереднене значення відносної помилки апроксимації тестової вибірки менше ніж в 0,03% для моделі з структурою 8-8-8 дає можливість вважати задовільним вибір як структури ІНС, так і навчального алгоритму, для розв'язання задачі апроксимації дискретного перетворення Хаара. Абсолютне усереднене значення помилки для всіх прикладів тестової вибірки і окремо для прикладів, не входивших в навчальну вибірку, відображено на графіку (рис. 3).

Модель ІНС навчається давати результати, які їй надані в навчальній вибірці. За рахунок здатності до узагальнення мережа може бути отримана нові результати, якщо подавати на вхід вектор, який не зустрічався при навчанні.

Для отримання результату по одному входному вектору ІНС з 8 нейронами в прихованому шарі, модель якої була сгенерована, необхідно виконати около 20 000 простіших операцій з плаваючою точкою. З урахуванням того, що швидкість сучасних ПК становить кілька сотень Mflops, даний розрахунок буде виконаний на ПК менше ніж за 1 сек. Навчання – більш витратна по часу процедура і може зайняти кілька десятків хвилин.

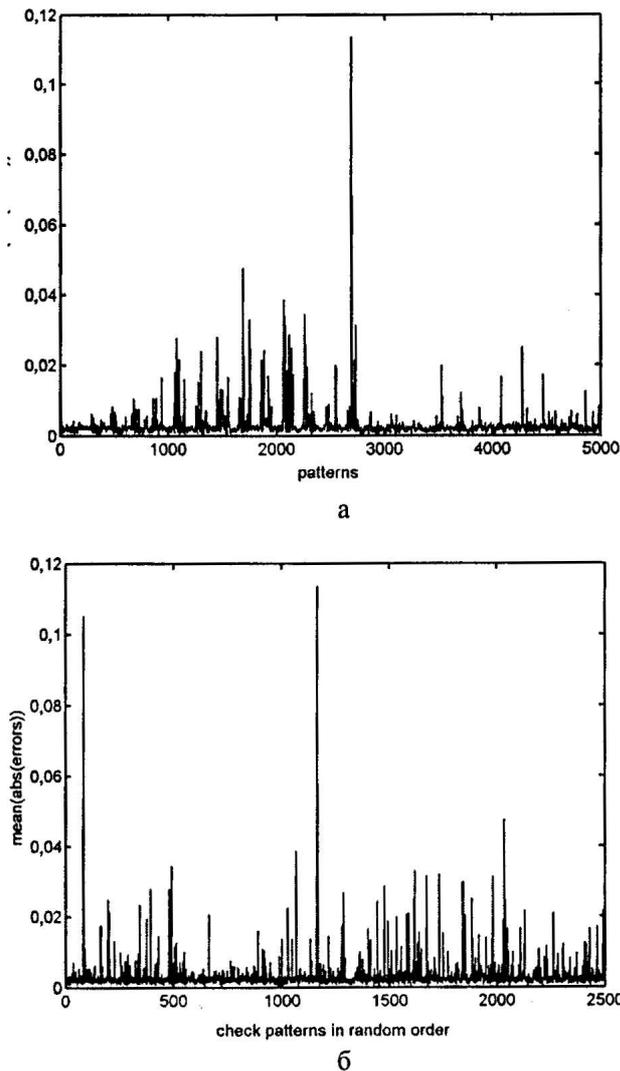


Рис. 3. Усредненная абсолютная ошибка аппроксимации моделью 8-8-8:

а – значения для всей тестовой выборки;  
б – значения для примеров, не вошедших в тестовую выборку (расположены в случайном порядке)

### Заключение

Рассмотренная среда моделирования, как и все ИНС, обладает уникальным свойством – универсальностью. Хотя для решения почти всех задач функционального преобразования существуют эффективные математические методы решения, и ИНС проигрывают специализированным методам для решения конкретных задач, благодаря свойству универсальности они становятся важным направлением исследования. Мы можем переучить уже имеющуюся модель для аппроксимации другого функционального преобразования той же размерно-

сти. Достаточно только выбрать корректный метод подготовки и формирования новой обучающей выборки.

Улучшить точность аппроксимации без увеличения размерности используемой структуры можно при помощи другого способа подготовки данных. Мы можем воспользоваться нелинейными преобразованиями. С учетом диапазона и функции распределения данных возможно использование логарифмической шкалы для подготовки данных.

### Литература

1. Претт У. Цифровая обработка изображений. Кн. 2. – М.: Мир, 1982. – 480 с.
2. Иванов В.Г. Формальное описание дискретных преобразований Хаара // Проблемы управления и информатики. – 2003. – № 5. – С. 35 – 40.
3. Суворовцев И.С., Клюкин В.И., Пивоварова Р.П. Нейронные сети. – Воронеж: ВГУ, 1994. – 224 с.
4. Bernard Widrow, Michael A. Lehr. 30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation // Artificial Neural Networks: Concepts and Theory, IEEE Computer Society Press. – 1992. – P. 327 – 354.
5. Нейроинформатика / А.Н. Горбань, В.Л. Дунин-Барковский, А.Н. Кирдин и др. – Новосибирск: Наука, Сибирское предприятие РАН, 1998. – 296 с.
6. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. – М.: Мир, 1985. – 509 с.
7. Hornik K. Approximation Capabilities of Multilayer Feedforward Networks // Neural Networks. – 1991. – № 4. – P. 251 – 257.
8. Галушкин А.И. Теория нейронных сетей. Кн.1: Учебное пособие для вузов – М.: ИПРЖР, 2000. – 416 с.

Поступила в редакцию 25.11.2004

Рецензент: д-р. техн. наук, проф. А.Ю. Соколов, Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Харьков.