

# ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ТЕХНОЛОГИЙ ПЕРЕДАЧИ И ОБРАБОТКИ ИНФОРМАЦИИ И СИГНАЛОВ

УДК 519.688:004.032.26

## ИСПОЛЬЗОВАНИЕ ВЕЙВЛЕТОВ ДЛЯ УЛУЧШЕНИЯ ПАРАМЕТРОВ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ РАСПОЗНАВАНИЯ РЕЧИ

Червяков Н.И., Астапов К.А.

Предлагается новый способ улучшения нейронных сетей с помощью добавления модуля обратного вейвлет-разложения. Рассматривается возможность применения нейронных сетей с модулем обратного вейвлет-разложения для задач распознавания речи. Приводятся результаты численного моделирования на ЭВМ.

### Введение

Проблема интеллектуальности распознающих систем заключается в том, что хотя они уже превзошли человека в точности распознавания, но не могут использовать информацию контекста для более правильной оценки результата, в отличие от человека, для которого задача распознавания речи лишь небольшая часть процесса ее понимания.

Природа речевого сигнала такова, что задачу распознавания речи следует поставить особняком среди других задач распознавания. Прежде всего, следует отметить временную природу сигнала, что отражено в работах [1-4]. Исследователям приходится учитывать, что речевой сигнал псевдопериодичен, нестационарен и имеет части, зависящие от взаимоположения. Высокая временная вариативность или нестационарность речевого сигнала получила в английских источниках название «Time warping», или искажение времени [1; 4].

Для увеличения интеллектуальности систем автоматического распознавания многим исследователям видится перспективным использовать нейронные сети. Но сложность состоит в том, что классические ИНС плохо приспособлены к работе с сигналами временной природы и локальными искажениями масштаба. Специализированные же ИНС решают эту задачу, но временная вариативность, с которой они могут справиться, низка и ограничена масштабом отдельных звуков и слов.

Одним из способов увеличить эту вариативность является создание гибридных методов. Авторам работы представляется, что наиболее перспективно сочетание ИНС с вейвлет-анализом.

### Математическая постановка задачи и модель предлагаемой вычислительной системы

Математическая постановка задачи такова.

Дан ряд  $x = \{x_1, x_2, \dots, x_n\}$ , представляющий отсчеты некоторого сигнала (например, речевого), взятые в соответствии с теоремой Котельникова. Этому ряду сопоставлен ряд  $p = \{p_1, p_2, \dots, p_n\}$ , представляющий собой отсчеты целевой функции, где  $x(t)$  и  $p(t)$  – функции от переменной времени  $t$ . Требуется построить вычислительную систему, осуществляющую отображение  $x \rightarrow p$ .

В силу изложенных выше соображений, эту вычислительную систему авторам представляется целесообразным строить на основе нейронных сетей и вейвлетов.

Введем условные обозначения, как это показано на рис. 1. Обозначим  $x_i$  входные данные,  $p_i$  – желаемый результат,  $y_i$  – выход нейронной сети,  $q_i$  – выход системы (то есть результат обратного вейвлет-преобразования, произведенный над  $y_i$ ),  $E(w)$  – целевая (энергетическая) функция всей системы.

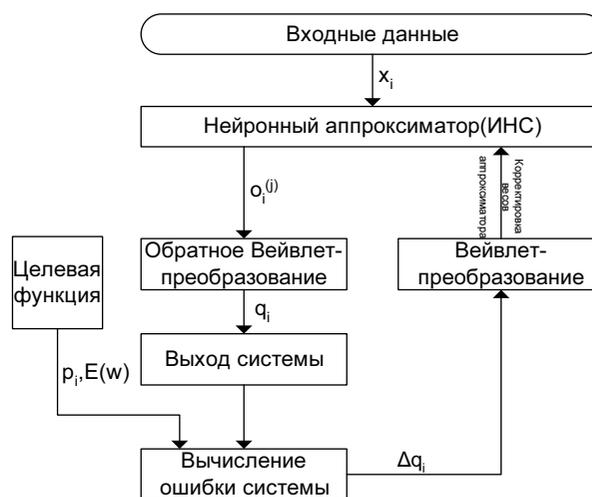


Рис. 1. Условные обозначения

В то же время обратное вейвлет-преобразование можно представить с коннекционистских позиций, как слой нейронной сети. В самом деле,

одноуровневая обратная вейвлет-декомпозиция может быть представлена как

$$y_i^{(l)} = \sum_{i_1=1}^N (h_{i-2i_1} a_{i_1}^{(l+1)} + g_{i-2i_1} d_{i_1}^{(l+1)}), \quad (1)$$

где  $a_{i_1}^{(l+1)}$  – «грубая» версия сигнала,  $d_{i_1}^{(l+1)}$  – детализирующие коэффициенты, а  $h_{i-2i_1}$  и  $g_{i-2i_1}$  – коэффициенты, характеризующие обратное вейвлет-преобразование.

При этом исходный сигнал может быть представлен последовательным применением формулы (1). Легко увидеть, что если заменить обозначения, введя единую нумерацию для  $a_{i_1}^{(l+1)}$  и  $d_{i_1}^{(l+1)}$ , обозначив эти значения как  $x_i$ , и так же ввести единые обозначения для  $a_{i_1}^{(l+1)}$  и  $d_{i_1}^{(l+1)}$ , обозначив их  $w_i$ , мы получим формулу формального  $y_i^{(l)} = \sum_{i=0}^{2N} w_i x_i$ , что формально соответствует формуле искусственного нейрона с линейной функцией активации [5].

Таким образом, модуль обратной вейвлет-декомпозиции можно представить в виде нейронной сети с числом слоев, равных уровню разложения.

На рис. 2 представлено одноуровневое обратное преобразование Хаара [7] как слой нейронной сети. Здесь круги со знаком «+» означают функцию суммирования, со знаком «-» вычитания.

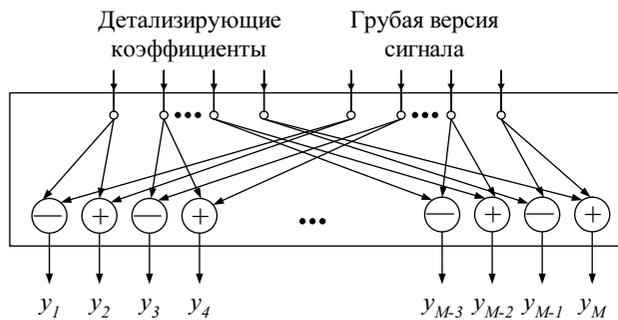


Рис. 2. Модуль одноуровневого обратного вейвлет-преобразования как слой нейронной сети

Таким образом, ИНС с модулем обратной вейвлет-декомпозиции может быть представлена как сложная нейронная сеть, в которой первые слои представляют собой некоторую стандартную нейронную сеть (например, многослойный перцептрон) и несколько слоев с линейными функциями активации, производящими обратное вейвлет-преобразование.

Пример такой нейронной сети, представляющей собой двухслойный перцептрон с модулем

обратного преобразования Хаара, можно видеть на рис. 3.

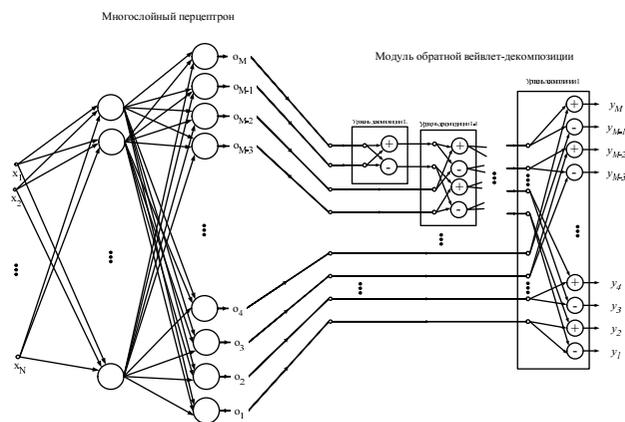


Рис. 3. ИНС из двухслойного перцептрона с модулем обратной вейвлет-декомпозиции Хаара

### Использование для задач распознавания речи

Основная идея использования нейронных сетей с модулем обратной вейвлет-декомпозиции для задач распознавания речи заключается в сокращении области значений нейронной сети. В предыдущем разделе мы ввели целевую функцию как вероятность принадлежности данного участка распознаваемому звуку. Можно с уверенностью утверждать, что эта функция имеет большую степень гладкости, то есть что разность соседних отсчетов будет много меньше максимального значения, принимаемого активационной функцией выходного слоя:

$$\|p_{2i} - p_{2i-1}\| \ll M, \quad (2)$$

где  $M$  – амплитуда значений выходной активационной функции, а  $i = 1, 2, 3, \dots$

Практические эксперименты показывают, что  $\|p_{2i} - p_{2i-1}\|$  гарантированно меньше  $M$  в 3-5 раз. Это позволяет найти такой коэффициент  $k$ , что  $k\|p_{2i} - p_{2i-1}\| \leq M$ . Применяя одноуровневое преобразование Хаара, мы получаем

$$\begin{aligned} \tilde{p}_{2i} &= 0,5 \|p_{2i} - p_{2i-1}\|, \\ \tilde{p}_{2i-1} &= 0,5 \|p_{2i} + p_{2i-1}\|; \end{aligned} \quad (3)$$

Таким образом, зная, что  $\tilde{p}_{2i-1}$  принимает ограниченные значения, мы можем без всякого риска умножить его на  $k$ . В виду ограниченности активационной функции эта операция сократит область поиска в  $k$  раз. Так, если мы используем биполярную активационную функцию с облас-

тью значений  $(-1,1)$ , а значения  $\tilde{p}_{2i-1} \in \left(-\frac{1}{5}; \frac{1}{5}\right)$ , то умножение на 5 позволит избежать поиска нейронной сетью решений на отрезках  $\left[-1; -\frac{1}{5}\right] \cup \left[\frac{1}{5}; 1\right]$  и сократить область значений в 5 раз. В практических экспериментах был использован размер выходного вектора равный 16. Возможность отмасштабировать каждый второй компонент вектора хотя бы в три раза уменьшает область значений (улучшает локализацию решения) в 6561 раз, в пять – в 390.625 раз! Таким образом, мы получаем возможность учитывать априорную информацию о целевом векторе за счет линейного преобразования выходов, увеличивая стабильность и скорость работы нейронной сети за счет уменьшения зоны поиска и потери универсальности. При таком подходе нейронная сеть может аппроксимировать лишь узкий класс векторов, но, как показывают описанные ниже эксперименты, качество работы и скорость сходимости улучшается на порядки.

### Обучение нейронной сети с модулем обратной вейвлет-декомпозиции

Пусть модуль обратной вейвлет-декомпозиции осуществляет преобразование, обратное вейвлет-преобразованию, заданному двумя зеркально-квадратурными FIR фильтрами G и H, определяемыми коэффициентами  $h_i$  и  $g_i$ .

Тогда для ИНС с модулем обратной декомпозиции имеют место формулы обучения, сходные с формулами метода обратного распространения ошибки:

$$\delta_{I(i,j)}^N = (q_i - p_i) K_i^{(j)} \frac{do_i^{(j)}}{d\left(\sum_{k=0}^{n_t} w_{kl(i,j)}^N x_{kj}^N\right)}, \quad (4)$$

$$\delta_i^t = \left(\sum_{l=1}^{n_t} \delta_l^{t+1} w_{il}^{t+1}\right) \frac{dy_i^t}{d\left(\sum_{k=0}^{n_t} w_{ki}^t x_{kj}^t\right)}, \quad (5)$$

$$\Delta w_{ij}^t = -\alpha \cdot \delta_j^t \cdot x_{ij}^t. \quad (6)$$

Здесь  $o_i^{(j)}$  – выходные значения ИНС (вейвлет-коэффициенты), индекс  $j$  означает уровень вейвлет-разложения,  $i$  – номер индекса на данном уровне; формальная функция  $I(i, j)$  сопоставляет двумерному индексу  $(i, j)$  одномерный; коэффициент  $K_i^{(j)} = \frac{\partial q_j}{\partial o_i^{(j)}}$

$$K_i^{(j)} = \frac{1}{\sum_{i_1=1}^{n_H} \sum_{i_2=1}^{n_H} \dots \sum_{i_{j-2}=1}^{n_H} \left(\prod_{r=1}^{r=j-1} h_{i_r-2i}\right)},$$

для  $j = N_I$ ; (7)

$$K_i^{(j)} = \frac{1}{\sum_{i_1=1}^{n_G} \sum_{i_2=1}^{n_H} \dots \sum_{i_{j-2}=1}^{n_H} \left(g_{i_1-2i} \prod_{r=2}^{r=l-1} h_{i_r-2i}\right)},$$

для  $j \in [1; N_I)$ ; (8)

$q_j$  – выходное значение системы;  $w_{ij}^t$  –  $j$ -ый вес  $i$ -го нейрона слоя  $t$ ;  $x_{i-}^t$  –  $i$ -ое входное значение слоя  $t$ ;  $y_i^t$  –  $i$ -ое выходное значение слоя  $t$ ;  $\alpha$  – коэффициент скорости обучения.

Из этих формул мы видим, что в результате применения разложения целевой функции по временному базису происходит ранжирование выходов нейронной сети по значимости и каждый вход обучается со своей собственной скоростью, соответствующей его значимости в вычислении конечного результата, что способствует повышению устойчивости и скорости процесса в целом.

### Результаты экспериментов

Тестирование и экспериментальная проверка нейронной сети с модулем обратной вейвлет-декомпозиции проводилась на задаче распознавания изолированного звука.

Для экспериментов был выбран речевой корпус из 37 слов. Проверка нейронной сети с модулем обратного вейвлет-разложения проводилась на задаче распознавания изолированного звука «а». Задача усложнялась тем, что длина звука в базе варьировалась в пределах четырехкратной от минимальной.

Целью эксперимента являлось сравнение двух абсолютно одинаковых нейронных сетей, с идентичной архитектурой, обучаемых одним и тем же алгоритмом, одна из которых обучалась с модулем вейвлет-разложения цели, а другая без него.

Оценка нейронных сетей проводилась следующим образом. Было проведено несколько запусков системы (здесь приводятся результаты для 20 запусков). В каждом запуске был выбран контрольный пример с худшим результатом и вычислено математическое ожидание ошибки по всем примерам контрольной выборки. По результатам двадцати запусков был выбран лучший результат по обоим критериям.

В ходе описанных экспериментов были получены следующие результаты (см. таблицу 1).

Таблица 1. Результаты экспериментов

Метод	Критерий №1, %	Критерий №2, %	Число эпох	Дисперсия
Перцептрон	11,57	17,83	3873	0,03028
Перцептрон с вейвлет-модулем	0,00330	0,00744	385	0,00016

Здесь «критерий №1» – это лучшее математическое ожидание ошибки на контрольной выборке по результатам обучения 20 нейронных сетей. Соответственно «критерий №2» – это лучшее математическое ожидание супремума ошибки на контрольной выборке по результатам обучения 20 нейронных сетей.

По результатам экспериментов можно говорить о примечательных результатах – которые, как это видно по дисперсии, весьма устойчивы. Можно также отметить резкое ускорение сходимости и качественное возрастание обобщающих способностей.

В таблице 2 приводятся дисперсия и математическое ожидание ошибки для эксперимента.

Таблица 2. Статистические характеристики

Метод	Математическое ожидание	Дисперсия
Перцептрон	15,50	0,03028
Перцептрон с вейвлет-модулем	0,014	0,00016

Из таблицы мы видим, что модуль вейвлет-разложения позволяет получить результат с небольшим математическим ожиданием и смещением ошибки, что говорит о хорошем качестве решения и адаптивности метода [5].

## Выводы

Получена математическая модель нового типа гибридов нейронных сетей и вейвлетов, предназначенных для повышения качества нейронных сетей, применимая к задачам распознавания речи. Теоретически обоснованы преимущества использования модели, выведены рекурсивные формулы, аналогичные алгоритму распространения ошибки и предложен алгоритм обучения для данной математической модели.

В результате практических вычислительных экспериментов получены результаты, значительно превосходящие стандартные, с малым смещением и дисперсией функции ошибки. С помощью предлагаемой схемы результат был улучшен на три-четыре порядка, что позволяет говорить о повышении качества распознавания для высоковариативных сигналов, таких как речевые.

## Литература

1. Tebelskis J. Speech Recognition using Neural Networks: PhD thesis ... Doctor of Philosophy in Computer Science. School of Computer Science, Carnegie Mellon University. Pittsburgh, Pennsylvania, 1995. – 179 p.
2. Lakhmi C., Jain L.C., Martin N.M. Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications. CRC Press, CRC Press LLC, 1998. – 297 p.
3. Handbook of neural network signal processing // Edited by Yu Hen Hu, Jenq-Neng Hwang. Boca Raton; London; New York, Washington D.C.: CRC press, 2001. – 384 p.
4. Principe J.C. Artificial Neural Networks. The Electrical Engineering Handbook // Ed. Richard C. Dorf. Boca Raton: CRC Press LLC, 2000. – 2719 p.
5. Хайкин С. Нейронные сети: полный курс. Пер. с англ. М.: ООО «ИД Вильямс», 2006. – 1104 с.
6. Добеши И. Десять лекций по вейвлетам. Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. – 464 с.
7. Haar A. Zur Theorie der Orthogonalen Funktionen-Systeme // Math. Ann. No. 69, 1910. – P. 331-371.

УДК 621.396.4

## ВЛИЯНИЕ КВАНТОВАНИЯ КОЭФФИЦИЕНТОВ ПЕРЕДАТОЧНОЙ ФУНКЦИИ ЦИФРОВОГО ФИЛЬТРА НА ЕГО ХАРАКТЕРИСТИКИ

*Тихобаев В.Г., Рясный Ю.В. Панарин В.И.*

Рассмотрена методика определения числа разрядов двоичного кода при квантовании коэффициентов передаточной функции цифрового фильтра, обеспечивающего заданные требования к точности амплитудно-частотной характеристики. Приведена методика вычисления абсолютной погрешности

импульсной характеристики, появляющейся вследствие ограничения числа разрядов.

## Введение

При реализации цифровых фильтров значения коэффициентов передаточной функции квантуются