

УДК 004.352.242

## РАЗРАБОТКА СИСТЕМЫ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ СИМВОЛОВ НА ОСНОВЕ СОВМЕСТНОГО ПРИМЕНЕНИЯ ВЕРОЯТНОСТНОЙ НЕЙРОННОЙ СЕТИ И ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

П.А. Хаустов, Д.С. Григорьев, В.Г. Спицын

Томский политехнический университет  
E-mail: eXceibot@sibmail.com; \_tryGX@sibmail.com.

Существует множество подходов к решению задачи оптического распознавания символов. Одним из них является нейросетевой метод распознавания, в котором для классификации используется нейронная сеть с топологией многослойного перцептрона. Нами разработан и представлен метод для анализа и классификации символов на основе применения вейвлет-преобразования для сокращения пространства признаков и вероятностной нейронной сети. Преимуществом вероятностной нейронной сети как классификатора является возможность использования вероятностного смысла выходных значений сети для улучшения качества распознавания. Осуществлен выбор оптимальной конфигурации параметров метода. Проанализированы результаты численных экспериментов по распознаванию символов, и проведена оценка быстродействия метода. В качестве исходных данных для тестирования предложенного метода использовались наборы символов с наличием пиксельного шума. Предложенный метод показал приемлемые результаты в 98 % правильно классифицированных символов.

### Ключевые слова:

Оптическое распознавание символов, вероятностная нейронная сеть, вероятностные оценки, вейвлет-преобразование, вейвлеты Хаара.

### Введение

Оптическое распознавание символов представляет собой механический или электронный перевод изображений рукописного, машинописного или печатного текста в текстовые данные – последовательность кодов, использующихся для представления символов в вычислительных системах. Алгоритмы оптического распознавания символов широко используются для конвертации книг и документов в электронный вид, для автоматизации систем учёта в бизнесе или для публикации текста на веб-страницах.

Распознавание латинских символов в печатном тексте с высокой степенью точности в настоящее время возможно лишь в случае работы с чёткими изображениями, такими как сканированные печатные документы. Точность при такой постановке задачи превышает 99 %, стопроцентная точность может быть достигнута лишь путем последующего редактирования человеком.

Достаточно часто для решения задачи оптического распознавания символов используют искусственные нейронные сети, которые обладают высоким быстродействием. Однако информация о классе, который искусственная нейронная сеть считает наиболее подходящим для заданного объекта, не всегда является достаточной для решения той или иной задачи. Иногда возможны ситуации, при которых один объект может быть в равной мере отнесен к двум и более классам. В таких случаях возникает необходимость в классификаторе, который позволит определять для каждого объекта некоторую степень принадлежности каждому из классов, аналогичную функции принадлежности в теории нечетких множеств. Для задачи оптического распознавания символов подобного рода информацию можно использовать для дальнейшего анализа результатов распознавания в системе оптического распознавания текстов.

Целью данной работы является разработка метода оптического распознавания символов, позволяющего получить оценку степени уверенности в результатах распознавания, а также определение оптимального набора дескрипторов для минимизации пространства признаков.

### Исходные данные

Исходные данные, которые были использованы при апробации метода оптического распознавания символов, представляют собой набор изображений отсканированных документов с присутствием пиксельного шума.

Примерная классификация типов пиксельного шума на изображениях приведена в таблице.

### Вероятностная нейронная сеть

Существенным преимуществом нейронных сетей является их высокое быстродействие, поэтому хорошим решением задачи распознавания символов являлась бы нейронная сеть, способная определять степень принадлежности образа к каждому из классов.

Одним из уже существующих решений является вероятностная модификация RBF-сети, которая носит название вероятностной нейронной сети (PNN-сеть) [1, 2]. По своей сути искусственная нейронная сеть такого типа решает задачу оценки плотности вероятности по имеющимся данным. Решение такой задачи в данном случае основано на так называемых ядерных оценках.

Фактически делается предположение о том, что существование наблюдения в некоторой точке пространства обеспечивает некоторую плотность вероятности в этой точке. Кластеры из близко лежащих точек указывают на то, что в этом месте плотность вероятности достаточно велика. По расстоянию от ядерных центров оценивается доверие к уровню плотности вероятности. Для оценки общей плотно-

сти вероятности используется суммарное значение некоторой функции во всех точках наблюдения. Такую функцию называют ядерной функцией. Зачастую в качестве ядерной функции используют распределение Гаусса. При больших размерах обучающей выборки такой метод дает достаточно точное приближение к истинной плотности вероятности.

Вероятностная нейронная сеть имеет три слоя: входной, радиальный и выходной. Входной слой имеет произвольную размерность. Каждый нейрон радиального слоя соответствует одному элементу обучающей выборки. Количество нейронов выходного слоя равняется количеству классов. Каждый нейрон выходного слоя соединен с элементами радиального слоя, принадлежащими к соответствующему ему классу.

Таким образом, для получения значения нейрона выходного слоя достаточно сложить отклики нейронов радиального слоя, с которыми он соединен. Для PNN-сети выход, соответствующий некоторому классу, описывается выражением (1):

$$y = \frac{1}{N\sigma^n} \sum_{k=1}^N \varphi\left(\frac{\|X - X^k\|}{\sigma}\right), \quad (1)$$

где  $n$  – размерность входного вектора,  $N$  – размер обучающей выборки,  $\sigma$  – параметр сглаживания,  $\varphi$  – Гауссова функция.

Выбор параметра сглаживания  $\sigma$  осуществляется отдельно для каждой задачи. Слишком большие значения  $\sigma$  приведут к потере деталей, слишком маленькие, наоборот, приведут к большому обобщению.

Таким образом, преимуществом PNN-сети является высокое быстродействие и вероятностный смысл значений выходных нейронов. Недостатком является зависимость количества нейронов скрытого слоя от размерности обучающей выборки [3, 4].

Сеть была апробирована на benchmark-наборе данных proben1, в результате чего было отмечено, что качество распознавания PNN-сети паритетно с качеством распознавания общеизвестных нейросетевых алгоритмов.

Обучение PNN-сети осуществляется с высокой степенью быстродействия. Ведь PNN-сеть не нуждается в обучении, достаточно лишь задать структуру радиального слоя этой сети.

Стоит отметить, что еще одним преимуществом PNN-сети является ее детерминированность. Результат распознавания при фиксированной обучающей выборке и фиксированном параметре  $\sigma$  всегда одинаков.

В результате проведенного исследования можно утверждать, что вероятностная нейронная сеть является хорошим инструментом для решения задачи классификации. Процент правильно выполненной классификации ничуть не уступает остальным вариациям нейронных сетей, однако появляется возможность использовать вероятностный смысл выходных значений для улучшения качества классификации или использования PNN-сети в качестве одного из классификаторов при комбинированном методе распознавания.

Таблица. Примерная классификация типов пиксельного шума

№	Изображение	Значение	Описание
1		a	Шум, вызывающий ошибки сегментации. Представляет собой отдельный набор пикселей, расположенных на небольшом (6–10 пикс.) расстоянии от символа. Правильность распознавания почти всегда обеспечивает работа словаря, так как остальные символы в слове могут быть сегментированы и распознаны сетью верно
		h	
		k	
		n	
		s	
2		d	Шум, представляющий собой пятна, перекрывающие символ, размерами до 50 % площади самого символа. Приемлемый процент распознавания достигается совокупностью работы сети с дескрипторами и словаря
		t	
		o	
		t	
		5	
3		p	Шум, представляющий собой пятна с площадью 1–8 пикселей, частично перекрывающий символ. Практически во всех случаях удается распознать символ. Наиболее часто встречающийся вид пиксельного шума
		m	
		e	
		t	
		t	
4		u	Шум, представляющий собой пятна площадью более половины площади символа. Символы практически не распознаются. Часто оказываются зашумлены и соседние символы. Возможность распознать слово, а соответственно и символ зависит от длины слова и количества загрязненных соседних символов
		r	
		g	
		e	
		t	
5		y	Шум, представляющий собой отсутствие составляющих символ пикселей. Практически всегда удается распознать символ
		e	
		i	
		l	
		e	

Очевидно, задача оптического распознавания символов является частным случаем задачи классификации образов, применения вероятностной нейронной сети для которой уже было описано ранее. Остается лишь определиться с пространством признаков, которые будут подаваться на входной слой такой сети.

Пусть все изображения символов, которые требуется классифицировать, являются бинарными изображениями фиксированного размера  $32 \times 32$  пикселей. В таком случае можно передавать на входы вероятностной нейронной сети значения яркости каждого из пикселей, которые всегда будут равны нулю или единице. Для таких дескрипторов изображения размерность пространства будет равна количеству бит в изображении (для указанных размеров это значение равно 1024).

Фактически при таком представлении каждое изображение будет являться некоторой вершиной 1024-мерного гиперкуба. В таком случае можно утверждать, что расстояние будет пропорционально количеству бит, которые в двух сравниваемых изображениях имеют различный цвет. То есть при таком представлении Евклидова метрика фактически идентична метрике манхэттенской с точностью до константы, на которую производится умножение.

Таким образом, при достаточных размерах обучающей выборки вероятностная нейронная сеть будет находить множество вершин гиперкуба, принадлежащих к одному и тому же классу, суммарное расстояние до которого в нем минимально.

Существенным недостатком такого подхода является достаточно большая размерность пространства признаков.

#### Формирование пространства признаков меньшей размерности

Вейвлеты являются приемлемым инструментом для обработки изображений и сигналов, позволяющим проводить анализ в нескольких временных масштабах локальных свойств. В частности, при помощи вейвлет-преобразования осуществляется сжатие. При оптимальном выборе базисной вейвлет-функции и уровня преобразования можно свести потерю качества к минимуму [5–7].

Применяемая в данной работе вероятностная нейронная сеть взаимодействует с набором вейвлет-дескрипторов и с матрицей вариаций яркости. Одной из целей данной работы является определение оптимального набора дескрипторов для минимизации пространства признаков.

В работе для получения коэффициентов разложения исходного изображения выбирается метод кратномасштабного анализа. Как известно, кратномасштабный анализ инвариантен к целочисленным сдвигам по временной оси и по отношению к растяжениям, кратным степеням двойки.

#### Получение дескрипторов

Так как цифровое изображение представляет собой функцию двух аргументов  $f(x, y)$ , то задав двумерные разделимые масштабирующую функцию и вейвлет-функции, обобщим алгоритм быстрого вейвлет-преобразования на двумерный случай.

Семейство базисных функций, определенное с помощью операций сдвигов и изменений масштаба, имеет вид:

$$\varphi_{j,m,n}(x, y) = 2^{j/2} (2^j x - m, 2^j y - n), \quad (2)$$

$$\psi_{j,m,n}^i(x, y) = 2^{j/2} \psi^i(2^j x - m, 2^j y - n). \quad (3)$$

$$i = \{H, D, V\}. \quad (4)$$

Индекс  $i$  в выражениях (2) и (3) служит для идентификации направленных вейвлетов, измеряющих вариации значений функции – изменения яркости для изображения по разным направлениям: вдоль столбцов, вдоль строк и вдоль диагоналей.

Определим дискретное вейвлет-преобразование (ДВП) изображения – функции  $f(x, y)$  для изображений размера  $M \times N$ . Аналогично разложению в ряд Фурье разложение в вейвлет-ряд ставит в соответствие непрерывной функции последовательность коэффициентов. В нашем случае разложение определяется следующими преобразованиями:

$$W_\varphi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \varphi_{j_0, m, n}(x, y),$$

$$W_\psi^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j, m, n}^i(x),$$

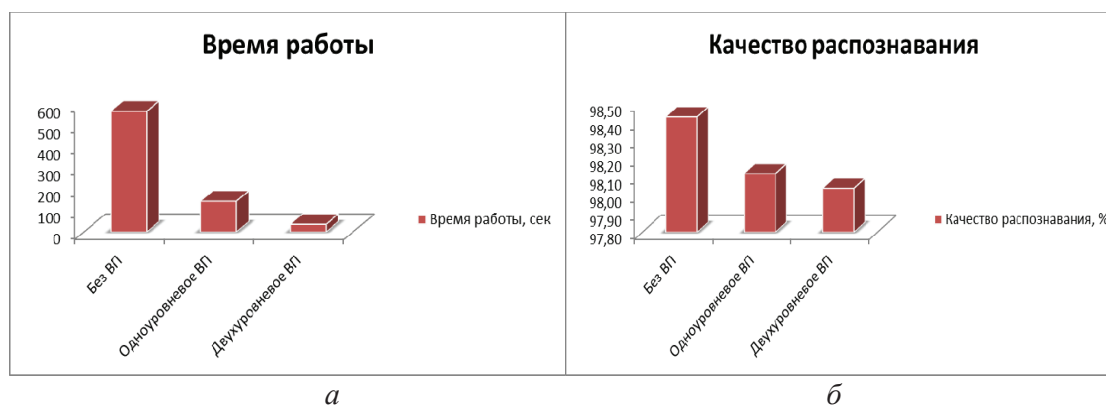
$$i = \{H, D, V\}.$$

Выражения (3) и (4) являются соответственно коэффициентами аппроксимации и коэффициентами детализации, где  $j_0$  – приближенный начальный масштаб, и коэффициенты  $W_\varphi(j_0, m, n)$  определяют приближение функции  $f(x, y)$  в масштабе  $j_0$ . Коэффициенты  $W_\psi^i(j_0, m, n)$  – горизонтальные, вертикальные и диагональные коэффициенты детализации для масштабов, больших чем  $j_0$ .

ДВП в нашем случае реализуется с помощью частотной фильтрации и прореживающей выборки. При этом происходит вычисление одномерного быстрого вейвлет-преобразования (БВП) по строкам функции, а затем по столбцам от уже полученного результата [6].

Тип используемой вейвлет-функции – функция Хаара. Порядок функции обусловлен величиной вектора коэффициентов аппроксимации. При его увеличении соответственно растет вектор коэффициентов, поданный на входной слой нейронной сети, что ведет к возрастанию вычислительной сложности.

После применения ДВП извлекались соответственно коэффициенты аппроксимации и коэффициенты детализации. В дальнейшем использовался только набор коэффициентов аппроксимации: в качестве точки отсчета для последующего применения блока фильтров БВП, а также в качестве входного сигнала для нейронной сети. Результат тестирования показал, что оптимальное количество применения блоков БВП фильтров равно 2. При каждом последующем применении блока БВП размер вектора коэффициентов аппроксимации сокращается вдвое, при уменьшающейся информативности самого вектора, а также при понижении точности распознавания.



**Рисунок.** а) Время работы метода с различными наборами дескрипторов; б) точность распознавания с различными наборами дескрипторов

В итоге каждый этап фильтрации дает на выходе четыре изображения, а точнее четыре «части» исходного изображения  $W_\varphi, W\psi^H, W\psi^V, W\psi^D$ , где каждое меньше исходного в 2 раза. Эти изображения представляют собой результат прореживающей выборки с фактором 2 по всем направлениям к массиву [7–9]. Элементами этого массива являются скалярные произведения изображения  $f(x,y)$  на двумерные масштабирующие и вейвлет-функции каждого соответствующего масштаба [8, 10].

#### Результаты тестирования метода

Тестирование предложенного метода осуществлялось на тестовой выборке из 48820 элементов, которая была составлена из изображений, полученных в результате сегментации набора отсканированных текстов с наличием пиксельного шума.

Тестирование производилось для трех методов получения дескрипторов: без использования вейвлет-преобразования (1024 дескриптора), с использованием одноуровневого вейвлет-преобразования (256 дескрипторов) и с использованием двухуровневого вейвлет-преобразования (64 дескриптора). Для каждого из методов была определена точность распознавания (в процентах от общего размера тестовой выборки) и время выполнения (в секундах). Результаты тестирования можно отобразить с помощью следующих диаграмм (рисунок).

Как можно заметить, существенное улучшение быстродействия достигается за счет использования вейвлет-преобразований. При использовании

одноуровневого вейвлет-преобразования время работы уменьшается примерно в 4 раза, при использовании двухуровневого вейвлет-преобразования – примерно в 16 раз. Такое уменьшение легко объясняется линейной зависимостью времени выполнения от размерности пространства признаков.

При таком существенном улучшении быстродействия и уменьшении размерности пространства признаков точность распознавания уменьшается не больше, чем на 0,4 %. Это говорит о том, что вейвлет-преобразование позволяет существенно улучшить быстродействие, ценой незначительного ухудшения качества распознавания.

#### Заключение

1. Предложено использование вероятностной нейронной сети для решения задачи оптического распознавания символов с целью последующего использования значений выходов нейронной сети для улучшения качества распознавания в системах оптического распознавания текстов.
2. Предложено применение дискретного вейвлет-преобразования для уменьшения размерности пространства признаков с целью улучшения быстродействия алгоритма распознавания.
3. В дальнейшем планируется использование предложенного метода в системе оптического распознавания текстов совместно с алгоритмами словарной работы и предобработки изображений.

*Работа выполнена при финансовой поддержке гранта РФФИ № 12-08-00296а.*

#### СПИСОК ЛИТЕРАТУРЫ

1. Круглов В.В., Дли М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. – М.: Физматлит, 2000. – 224 с.
2. Тархов Д.А. Нейронные сети. Модели и алгоритмы. – М.: Радиотехника, 2005. – 256 с.
3. Probabilistic Neural Networks. 2004. URL: [http://courses.cs.tamu.edu/rgutier/cpsc636\\_s10/specht1990pnn.pdf](http://courses.cs.tamu.edu/rgutier/cpsc636_s10/specht1990pnn.pdf) (дата обращения: 22.05.2013).
4. An Introduction to Probabilistic Neural Networks. 2003. URL: <http://www.psi.toronto.edu/~vincent/research/presentations/PNN.pdf> (дата обращения: 22.05.2013).
5. Misiti M., Misiti Y., Oppenheim G., Poggi J. Wavelets and their applications. – London: ISTE, 2007. – 352 с.

6. Блаттер К. Вейвлет-анализ. Основы теории. – М.: Техносфера, 2004. – 280 с.
7. Добеши И. Десять лекций по вейвлетам. – Ижевск: Регулярная и хаотическая динамика, 2001. – 464 с.
8. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2006. – 1072 с.
9. Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2002. – 344 с.
10. Хайкин С. Нейронные сети: полный курс. 2-е изд. – М.: И.Д. Вильямс, 2006. – 1104 с.

*Поступила 03.09.2013 г.*

UDC 004.352.242

## THE DEVELOPMENT OF OPTICAL CHARACTER RECOGNITION APPROACH ON THE BASIS OF JOINT APPLICATION OF PROBABILISTIC NEURAL NETWORK AND WAVELET TRANSFORM

P.A. Khaustov, D.S. Grigoryev, V.G. Spitsyn

Tomsk Polytechnic University

There are a lot of approaches to solve the problem of optical character recognition. One of them is the approach based on neural networks. The authors have proposed and implemented an approach of analyzing and classification based on wavelet transformation for reducing the feature space and probabilistic neural network for recognition. The main advantage of probabilistic neural network is the probabilistic significance of its output neurons which can be used to improve the quality of recognition. Optimal values of parameters were selected for such approach. The results of numerical experiments were analyzed and the time-performance of this approach was assessed. The set of noisy character images was used to assess the proposed approach. This method showed 98 % of acceptable recognition.

### Key words:

Optical character recognition, probabilistic neural network, probabilistic assessment, wavelet transform, wavelets.

### REFERENCES

1. Kruglov V.V., Dli M.I., R.Yu. Golunov. *Nechetkaya logika i iskusstvennyye neyronnye seti* [Fuzzy Logic and Artificial Neural Networks]. Moscow, Fizmatlit Publ., 2000. 224 p.
2. Tarkhov D.A. *Neyronnye seti. Modeli i algoritmy* [Neural Networks. Models and Algorithms]. Moscow, Radiotekhnika Publ., 2005. 256 p.
3. *Probabilistic Neural Networks*. 2004. Available at: [http://courses.cs.tamu.edu/rgutier/cpsc636\\_s10/specht1990pnn.pdf](http://courses.cs.tamu.edu/rgutier/cpsc636_s10/specht1990pnn.pdf) (accessed 22 May 2013).
4. *An Introduction to Probabilistic Neural Networks*. 2003. Available at: <http://www.psi.toronto.edu/~vincent/research/presentations/PNN.pdf> (accessed 22 May 2013).
5. Misiti M., Misiti Y., Oppenheim G., Poggi J. *Wavelets and their applications*. London, ISTE, 2007. 352 p.
6. Blatter K. *Veyvlet-analiz. Osnovy teorii* [Wavelet-analysis. The Main Theory]. Moscow, Tekhnosfera Publ., 2004. 280 p.
7. Dobeshi I. *Desyat lektiy po veyvletam* [Ten Lectures about Wavelets]. Izhevsk, Reguljarnaya i khaoticheskaya dinamika, 2001. 464 p.
8. Gonsales R., Vuds R. *Tsifrovaya obrabotka izobrazheniy* [Digital Image Processing]. Moscow, Tekhnosfera Publ., 2006. 1072 p.
9. Osovskiy S. *Neyronnye seti dlya obrabotki informatsii* [Neural Networks for Data Processing]. Moscow, Finansy i statistika Publ., 2002. 344 p.
10. Haykin S. *Neyronnye seti* [Neural Networks]. Moscow, Williams, 2006. 1104 p.