

Новая концепция нейросети для распознавания и классификации пиксельных образов

И.О. Гребенкин, Н.А. Магницкий, А.Ю. Чернявский

Аннотация. В работе предложен новый подход к решению задачи распознавания пиксельных черно-белых изображений с использованием искусственных нейронных сетей. Рассматривается алгоритм построения и обучения искусственной нейронной сети LICS (Linear Characters Separation), относящейся к многослойным персептронам или персептронам со скрытым слоем. Данный алгоритм имеет ряд преимуществ, среди которых – минимальное число устанавливаемых параметров и, как следствие, возможность работы алгоритма в целом классе задач без предварительной 'настройки' для каждой конкретной задачи, а также его наглядность и простота.

Ключевые слова: пиксельный образ, нейронная сеть, персептрон, алгоритм обучения.

1. Постановка задачи распознавания пиксельных образов

Пусть ставится задача распознать пиксельное изображение одного из K объектов (например, различных цифр), и для каждого из распознаваемых объектов задано некоторое множество M_k , $k = 1, \dots, K$, его возможных эталонных черно-белых изображений (написаний) в прямоугольном поле $r \cdot q = n$ пикселей. Тогда каждому изображению каждого объекта может быть поставлен в соответствие n – мерный входной вектор $x = (x_1, \dots, x_n)^T \in 2^n$, где $x_j = 1$, если j -ый пиксель принадлежит образу объекта, и $x_j = -1$, если j -ый пиксель не принадлежит образу объекта, $j = 1, \dots, n$. При этом k -ому объекту из K объектов соответствует M_k входных эталонных векторов из всего огромного множества всевозможных входных векторов 2^n . Общее количество эталонных входных векторов будет равно

$$M = \sum_{k=1}^K M_k.$$

2. Построение нейросети LICS

Для решения поставленной задачи построим искусственную нейронную сеть, формально являющуюся трехслойным персептроном, имеющим n входов, K выходов и N элементов в скрытом слое

$$g_k(x_1, \dots, x_n) = \sum_{i=1}^N \omega_{ki} \sigma_i \left(\sum_{j=1}^n \alpha_{ij} x_j - \beta_i \right),$$
$$k = 1, \dots, K. \quad (1)$$

Как известно, сеть (1) теоретически решает поставленную задачу, однако практически на сегодняшний день отсутствуют работающие алгоритмы обучения такой сети, т.е. алгоритмы выбора весов $\alpha_{ij}, \omega_{ki}, \beta_i$, $i = 1, \dots, N$; $j = 1, \dots, n$; $k = 1, \dots, K$. Так, например, метод *обратного распространения ошибки (backpropagation)* сводит проблему обучения сети (1) к нерешаемой задаче поиска глобального экстремума сложной нелинейной функции в пространстве всех весов [1]; *нейросеть Фурье* [2] сводит задачу к обучению однослойного персептрона с

весами ω_{ki} , однако при этом отсутствуют условия и оценки сходимости методов обучения, а также теоретические оценки величины N элементов в скрытом слое.

Предложим совершенно иной подход к построению и обучению нейросети (1), существенно использующий тот факт, что в задаче распознавания пиксельных образов входные векторы являются вершинами 2^n - мерного куба. Положим в (1) $N = M$, а векторы α_{ij}, ω_{ki} , $i = 1, \dots, N$; $j = 1, \dots, n$; $k = 1, \dots, K$, зададим следующим образом. Если $x_{mk}^* = (x_{mk1}^*, \dots, x_{mkn}^*)^T$ – входной вектор m -ого эталонного образа ($1 \leq m \leq M_k$) k -го объекта ($1 \leq k \leq K$), и

$$i = \sum_{s=1}^{k-1} M_s + m, 1 \leq m \leq M_k,$$

то положим $\alpha_{ij} = x_{mkj}^*$; $\omega_{ki} = 1$. Остальные веса ω_{ki} положим равными нулю. В качестве нелинейной функций скрытого слоя $\sigma_i(y)$ в (1) выберем ступенчатую функцию

$$\chi_i(y) = \chi_{mk}(y) = \chi(y - \beta_{mk}) = \begin{cases} 0, & y < \beta_{mk}, \\ 1, & y \geq \beta_{mk}. \end{cases}$$

Таким образом, нейросеть (1) примет вид

$$g_k(x_1, \dots, x_n) = \sum_{m=1}^{M_k} \chi \left(\sum_{j=1}^n x_{mkj}^* x_j - \beta_{mk} \right),$$

$$k = 1, \dots, K \quad (2)$$

Структурная схема нейросети LICS (2) представлена на Рис.1.

Нейросеть LICS имеет следующие преимущества перед другими нейронными сетями такого типа:

а) относительно небольшое и фиксированное число элементов (нейронов) в скрытом слое, равное общему числу M векторов обучающей выборки;

б) фиксированные связи $\alpha_{ij} = x_{mkj}^*$ между входными векторами и нейронами скрытого слоя, задаваемые координатами векторов эталонных образов;

в) фиксированные связи между нейронами скрытого слоя и нейронами выходного слоя, равные единице или нулю;

г) размерность пространства векторов, координаты β_{mk} которых необходимо определить в процессе обучения, равна относительно небольшому числу M всех эталонных образов.

3. Обучение нейросети LICS

Итак, в процессе обучения необходимо задать β_{mk} , $k = 1, \dots, K$; $m = 1, \dots, M_k$ – пороги ступенчатых функций χ_{mk} нейросети (2). Так как расстояние от нуля до любой вершины n -мерного куба равно $\sqrt{n}$, то, выбрав β_{mk} равным или немногим меньше n , то при подаче на вход сети вектора x_{mk}^* – m -го эталонного образа k -го объекта, получим в скрытом слое только χ_{mk} равную единице, а остальные χ равными нулю. Следовательно, на выходе будем иметь $g_k = 1$ и $g_p = 0$ при всех $p \neq k$. Таким образом, при таком выборе порогов β_{mk} , нейросеть (2) для каждого эталонного образа k -го объекта на k -ом выходе даст единицу, а на всех других выходах нули, т.е. все эталонные образы из обучающей выборки будут гарантированно правильно классифицированы.

Однако такая нейросеть будет обладать недостаточными аппроксимационными свойствами, т.е. она может давать ошибки на тести-

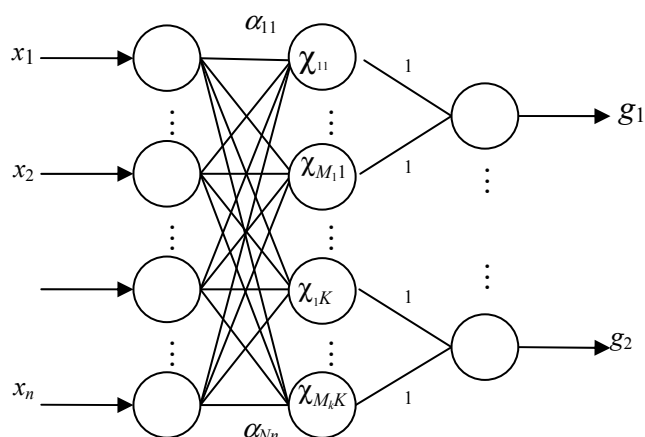


Рис. 1. Нейросеть LICS: трехслойный персептрон специального вида, состоящий из входного слоя, выходного слоя и одного промежуточного (скрытого) слоя из M элементов

рующей выборке. Чтобы уменьшить количество таких ошибок и, возможно, свести это количество к нулю, необходимо уменьшить значения порогов $\beta_{mk} < n$. При этом гиперплоскость

$$P_{mk}(\beta): \sum_{j=1}^n x_{mkj}^* x_j - \beta = 0$$

будет проходить на расстоянии $\beta/\sqrt{n}$ от центра n -мерного куба перпендикулярно эталонному вектору x_{mk}^* , отсекая от вершины, задаваемой вектором x_{mk}^* , некоторую область куба, содержащую не только саму вершину, но некоторое количество близких к ней вершин (Рис.2). При этом β определяет 'сдвиг' гиперплоскости.

Получилось, что теперь в скрытом слое выход x_{mk} будет равен 1 не только для входа x_{mk}^* , но также и для некоторого набора близких к нему входов. Соответственно, для всего этого набора входов получим на выходе сети $g_k \geq 1$ и $g_p = 0$ при всех $p \neq k$. Таким образом все близкие к эталонному вектору x_{mk}^* в указанном выше смысле векторы, задающие искаженный относительно эталонного образа k -ого объекта, будут также распознаны и классифицированы правильно, как k -ый объект.

Ясно, что чем меньше значение 'сдвига' β , тем лучше аппроксимационные свойства эталонного образа x_{mk}^* , т.е. тем больше его искаженных (похожих) образов (близких вершин куба) будут распознаны правильно. Однако существуют естественные границы уменьшения значений сдвига β . Этой границей является такое значение 'сдвига' β_{mk}^* , при котором в отсекаемую область куба попадает первая вершина, соответствующая некоторому эталонному образу некоторого отличного от k -го объекта, т.е. первая вершина x_{mp}^* , $p \neq k, 1 \leq m \leq M_p$, лежащая в гиперплоскости $P_{mk}(\beta_{mk}^*)$, для которой

$$\sum_{j=1}^n x_{mkj}^* x_{mpj}^* - \beta_{mk}^* = 0.$$

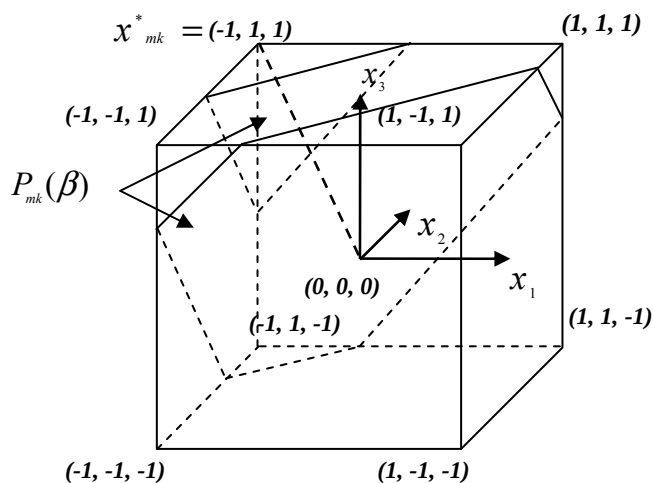


Рис.2. Пояснения к алгоритму обучения нейросети LICS для случая $n = 3$

Следовательно, естественная нижняя граница сдвига определяется формулой:

$$\beta_{mk}^* = \max_{p \neq k; 1 \leq m_p \leq M_p} \sum_{j=1}^n x_{mkj}^* x_{mpj}^* \quad (3)$$

Наилучший результат будет достигаться при:

$$\beta_{mk} = \beta_{mk}^* + \varepsilon, \quad (4)$$

где параметр ε – единственный параметр сети.

4. Результаты тестирования классификатора LICS на базе рукописных изображений MNIST

База MNIST является базой изображений рукописных цифр, содержащей 60000 обучающих и 10000 тестовых изображений. База находится в свободном доступе по веб-адресу <http://yann.lecun.com/exdb/mnist/>, по этому адресу можно получить подробное описание базы и результаты распознавания для некоторых алгоритмов. Изображения цифр в базе MNIST представляют собой массивы размером 28x28, каждый элемент которых имеет значение от 0 до 255 включительно и отвечает за яркость пикселя.

Тестирование классификатора LICS на базе MNIST можно разделить на три этапа: тестирование на чёрно-белых изображениях, тестирование на изображениях с градациями серого, тестирование с расширенной обучающей выборкой.

4.1. Тестирование на чёрно-белых изображениях

Входной вектор строится следующим образом: $x_i = \begin{cases} -1, & p_{i/28, i \bmod 28} = 0 \\ 1, & p_{i/28, i \bmod 28} > 0 \end{cases}$, где «/» - целочисленное деление, $\bmod j$ – остаток по модулю j , $p_{i,j}$ - яркость пикселя в точке (i, j) .

Как говорилось выше, единственным параметром классификатора является ε . На Рис.3 приводится график зависимости количества ошибок от этого параметра для задачи распознавания чёрно-белых изображений. Минимальное число ошибок достигается при $\varepsilon = 13$ и равно 370.

4.2. Тестирование на изображениях с градацией серого

Входной вектор строится следующим образом: $x_i = \frac{(p_{i/28, i \bmod 28} - 127.5)}{127.5}$, где «/» - целочисленное деление, $\bmod j$ – остаток по модулю j , $p_{i,j}$ - яркость пикселя в точке (i, j) .

Для снижения зависимости от параметра ε , а также для увеличения числа правильно распознанных тестовых изображений использовался следующий метод: если изображение не попало ни в один класс, то для него производится классификация методом ближайшего соседа, т.е. тестовый вектор x сравнивается со всеми векторами обучающей выборки y_i и среди них находится $y_{\max}$, для которого максимально $d(x, y_i) = \sum_j x_j y_{ij}$. Вектору x присваивается класс, соответствующий классу $y_{\max}$.

В результате тестирования были получены следующие результаты: без использования метода ближайшего соседа оптимальный параметр $\varepsilon = 11.6$, количество ошибок – 353, из них не попали ни в один класс - 44; с использованием метода ближайшего соседа оптимальный параметр $\varepsilon = 2.4$, количество ошибок – 325.

4.3. Тестирование с расширенной обучающей выборкой

Обучающая выборка была расширена за счёт циклических сдвигов исходных изображений

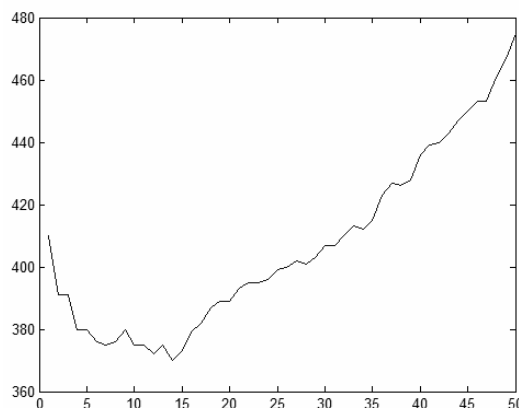


Рис.3. График зависимости количества ошибок от ε

на 1 пиксель во всех возможных 8-ми направлениях. Теперь каждому элементу $t(x, y)$ исходной обучающей выборки стали соответствовать 9 элементов

$s_{dx, dy}(x, y) = t((x + dx) \bmod 28, (y + dy) \bmod 28)$ расширенной обучающей выборки, где x, y - координаты пикселя, а dx, dy - целочисленные величины соответствующих сдвигов, принимающие значения в диапазоне $[-1, 1]$.

Для чёрно-белых изображений с расширенной обучающей выборкой получен наилучший результат – 299 ошибок, для изображений с оттенками серого – 260 ошибок.

Заключение

Предложена оригинальная нейронная сеть, специально адаптированная к решению задач распознавания пиксельных образов, полученных с помощью цифровых и веб-камер. Основными достоинствами сети являются: наглядность и простота, относительно небольшая размерность пространства векторов, координаты которых необходимо определить в процессе обучения, прямой метод обучения с помощью явной вычислительной процедуры и фактически единственный параметр сети, который требуется настроить на решение конкретной задачи. Апробация сети на решении сложной задачи распознавания рукописных цифр на базе MNIST (более 97% правильного распознавания) ставит ее в один ряд с лучшими на сегодняшний день сетями, опробованными на базе рукописных цифр, и демонстри-

рует ее хорошие потенциальные возможности в решении других сложных задач распознавания пиксельных образов.

Литература

1. Rumelhart D.E., Hinton G.E., Williams R.J. Learning internal representations by error propagation // In Parallel dis-

tributed processing. – Cambridge, MA: MIT Press, 1986, v.1, p.318-362.

2. Магницкий Н.А. Некоторые новые подходы к построению и обучению искусственных нейронных сетей // Нелинейная динамика и управление, вып.1, под ред. С.В.Емельянова и С.К.Коровина. – М.: Физматлит, 2001, с.87-98.

Гребенкин Илья Олегович. Студент кафедры НДС и ПУ факультета ВМК МГУ. Круг научных интересов: нейронные сети, математическое моделирование.

Магницкий Николай Александрович. Заведующий лабораторией Института системного анализа РАН. Окончил МГУ в 1974 году. Доктор физико-математических наук, профессор кафедры НДС и ПУ ф-та ВМК МГУ. Автор более 150 публикаций, в том числе 5 монографий. Область научных интересов: нелинейные дифференциальные уравнения, хаотическая динамика, нейронные сети, математическое моделирование. Тел. (499) 135-43-32.

Чернявский Андрей Юрьевич. Сотрудник фирмы «Бинейро», аспирант факультета ВМК МГУ. Автор 2-х публикаций. Область научных интересов: нейронные сети, математическое моделирование.