

Методика разработки нейросетевых алгоритмов для решения задач обработки изображений

(часть 1)

Наталья Томашевич (Москва)

В статье рассмотрены типовые ошибки, возникающие при разработке нейросетевых алгоритмов, предназначенных для решения задач обработки изображений. На базе описанной методики построения нейронной сети рассмотрены практические задачи: 1) распознавание туберкулезных бацилл и 2) выделение лиц на цифровых изображениях. Обсуждаются результаты вычислительных экспериментов.

ВВЕДЕНИЕ

Важной особенностью задачи обработки изображений являются требования инвариантности (нечувствительности). К системе обработки изображений могут предъявляться требования линейной инвариантности, которые возможно формализовать, – инвариантность к сдвигу, повороту, отображению наблюдаемого объекта; и требования нелинейной инвариантности, которую нельзя описать формулами, – инвариантность к освещению, к помехам на изображении (снег, дождь, ошибки экспозиции и т.п.), к нелинейным искажениям объекта (при распознавании самолётов – новые виды самолётов или другие ракурсы съёмки, при распознавании лиц – мимика, положение лица в пространстве, масштаб и т.д.).

К подавляющему большинству задач обработки изображений предъявляются требования нелинейной

инвариантности, удовлетворить которые возможно только с использованием нелинейных методов. Однако мы не будем рассматривать их недостатки и преимущества, а перейдём к наиболее универсальному, на наш взгляд, методу – нейронным сетям.

У людей, мало знакомых с нейронными сетями, может сложиться впечатление, что по количеству разновидностей нейронные сети не уступают всем остальным методам обработки изображений. Однако впечатление о безграничности видов нейронных сетей обманчиво. Ещё в 1960-е годы российской научной школой было сформулировано единое универсальное описание, позволяющее в одной структуре обобщить все виды нейронных сетей (см. рис. 1). Как видно из рисунка, здесь присутствуют не только последовательные связи, но и обратные и перекрёстные связи, наличие и расположение которых может варьироваться. В зависимости от задачи используется разное число слоёв, нейронов в слоях и т.д. Сами нейроны могут иметь различные функции активации: линейную, нелинейную, радиальные базисные функции, периодические функции (wavelet) и др.

По способу обучения нейронные сети делятся на нейронные сети с фиксированной структурой и нейронные сети с переменной структурой. В нейронной сети с фиксированной структурой число слоёв, нейронов и их связи определены до

начала обучения сети. И уже для заданной структуры сети проводится её обучение, т.е. настройка весов каким-либо, например, градиентным, методом. В нейронной сети с переменной структурой она заранее неизвестна, и сеть строится по данным обучающей выборки. К нейронным сетям с переменной структурой относятся, например, иерархические нейронные сети.

Решение задач обработки изображений в нейросетевом логическом базисе, как правило, проходит по следующему сценарию. В каждой точке (x, y) исходного изображения формируется окно размера $N \times N$ (для реализации инвариантности к сдвигу объекта на изображении). Из содержимого окна формируется вектор признаков, который подаётся на вход нейронной сети. На выходе нейронная сеть должна выдать +1, если окно содержит искомым объект (например, лицо человека или самолет), и -1, если не содержит.

Таким образом, задача обработки изображения сводится к задаче классификации с двумя классами образов. Для реализации инвариантности к масштабу, повороту и отражению искомого объекта в каждой точке (x, y) формируется не одно, а множество окон разного размера и ориентации, которые подаются на вход сети. Далее мы рассмотрим нейросетевую постановку задачи обработки изображений, а именно, задачу классификации в N -мерном пространстве признаков с двумя классами.

Более подробно мы остановимся на разнице между нейронными сетями с переменной и фиксированной структурой применительно к задачам обработки изображений. Как было сказано выше, мы рассматриваем лишь те задачи обработки изображений, которые в своей нейросетевой постановке сводятся к задаче класси-

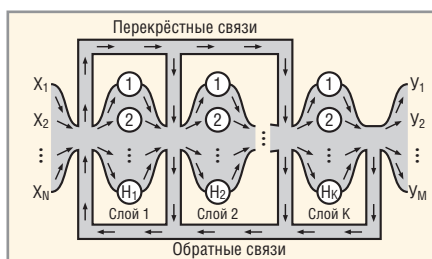


Рис. 1. Универсальная структура нейронной сети

Вектор $\{x_1, x_2, \dots, x_N\}$ – входной сигнал нейронной сети, вектор $\{y_1, y_2, \dots, y_M\}$ – выходной сигнал сети, H_i ($i = 1 \dots K$) – число нейронов в i -м слое сети, K – число слоёв сети.

фикации в многомерном пространстве признаков при числе классов образов, равном двум. То есть в многомерном пространстве признаков мы должны построить разделяющую поверхность, которая будет отделять один класс от другого. Для упрощения рассмотрим входное пространство, состоящее из двух признаков x_1 и x_2 .

Исследования многослойных нейронных сетей показывают, что в них каждый нейрон первого слоя реализует разделяющую гиперплоскость во входном пространстве признаков. Нейроны остальных слоёв лишь по-разному группируют эти гиперплоскости в области. В случае двумерного входного пространства каждый нейрон первого слоя сети будет реализовывать разделяющую прямую (см. рис. 2). Нейроны остальных слоёв объединяют эти разделяющие прямые в области, в результате чего и получается нелинейное разделение входного пространства.

В случае использования нейронной сети с фиксированной структурой мы берём фиксированное количество нейронов для каждого слоя. Обозначим число нейронов первого слоя Q1, второго слоя Q2, третьего слоя Q3 и т.д. Взяв какое-то число нейронов первого слоя Q1, мы по сути выбрали Q1 разделяющих прямых. Далее нам необходимо инициализировать веса нейронов какими-то значениями. Как правило, это делается случайным образом, в некотором небольшом интервале, например, от $-0,5$ до $+0,5$. Тогда все разделяющие прямые сгруппируются вокруг центра координат не далее

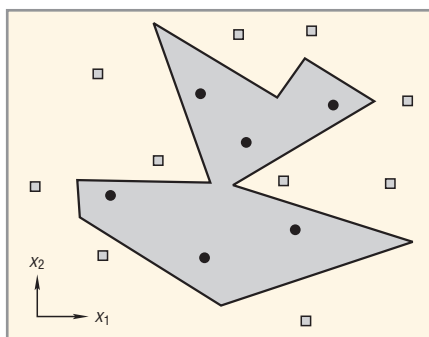


Рис. 2. Пример разделяющей области во входном пространстве двух признаков x_1 и x_2 с двумя классами образов

чем на 0,5. Нейроны остальных слоёв не влияют на расположение нашего пучка прямых, – в зависимости от значений весов они лишь по-разному объединяют эти прямые в области. Пример такой инициализации показан на рисунке 3. В принципе, имея для нейронной сети такую разделяющую поверхность (показана на рис. 3 тёмно-серым цветом), её можно обучить до получения правильной разделяющей поверхности (светло-серой). Но это при условии, что мы заранее выбрали правильно число слоёв, количество нейронов каждого слоя и правильно установили связи между ними. При решении практических задач это маловероятно.

Итак, для вышеуказанного класса задач обработки изображений более подходящими являются нейронные сети с переменной структурой. Наиболее широко в этой группе методов представлены иерархические нейронные сети. Перейдём к рассмотрению разработанной методики построения нейронной сети.

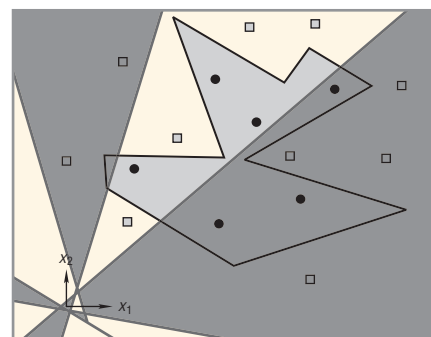


Рис. 3. Разделяющая поверхность, реализуемая нейронной сетью со случайным образом инициализированными значениями весов (тёмно-серая) по сравнению с правильной разделяющей поверхностью

МЕТОДИКА ПОСТРОЕНИЯ НЕЙРОННОЙ СЕТИ

1. Построение оптимальной гиперплоскости, разделяющей две точки i и j в N -мерном пространстве

Дано: (x_1^i, \dots, x_N^i) , (x_1^j, \dots, x_N^j) – координаты точек i и j в N -мерном пространстве. Точка i принадлежит классу A ($i = -1$), точка $j \in B$ ($j = +1$) (см. рис. 4а). Обозначим через d_{ij} расстояние между точками i и j , тогда:

$$d_{ij}^2 = \sum_{k=1}^N (x_k^j - x_k^i)^2.$$

Уравнение гиперплоскости, разделяющей точки i и j , выглядит следующим образом:

$$w_0 + \sum_{k=1}^N w_k x_k = 0. \quad (1)$$

Требуется: найти коэффициенты w_k ($k = 0..N$) гиперплоскости (1), оптимально разделяющей точки i и j

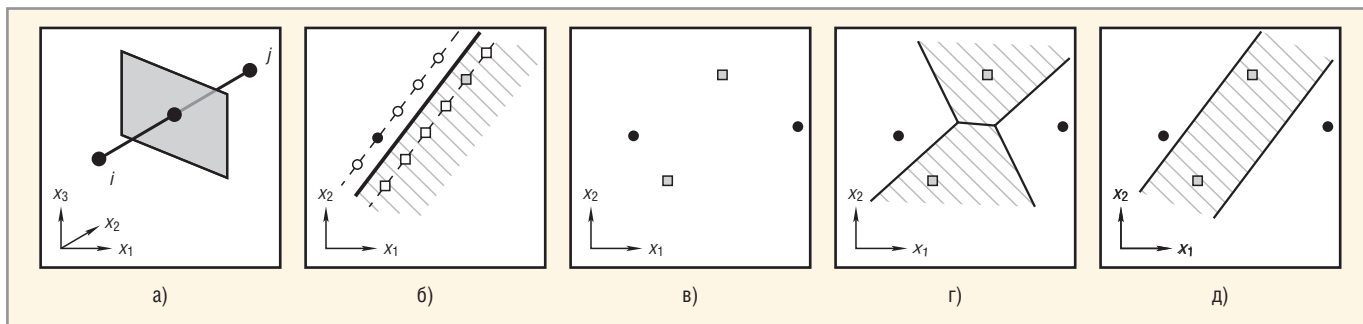


Рис. 4. Построение оптимальной гиперплоскости

- (а) Пример построения гиперплоскости (серый прямоугольник), разделяющей точки i и j в 3-мерном ($N = 3$) пространстве.
- (б) Пример построения гиперплоскости в двумерном пространстве с инвариантностью к яркости изображения. Чёрный кружок – точка i ; серый квадрат – точка j ; пунктиром показаны прямые, определяемые точками $i \in A$ и $j \in B$ в двумерном входном пространстве; белые кружки и квадратики соответствуют точкам классов A и B соответственно; жирная чёрная линия – оптимальная разделяющая поверхность между точками i и j .
- (в) Пример обучающей выборки. Два класса образов в двумерном входном пространстве: чёрные кружки соответствуют одному классу, серые квадратики – другому.
- (г) Пример построения оптимальной разделяющей плоскости для заданной обучающей выборки.
- (д) Пример построения оптимальной разделяющей плоскости с учётом инвариантности к яркости изображения.

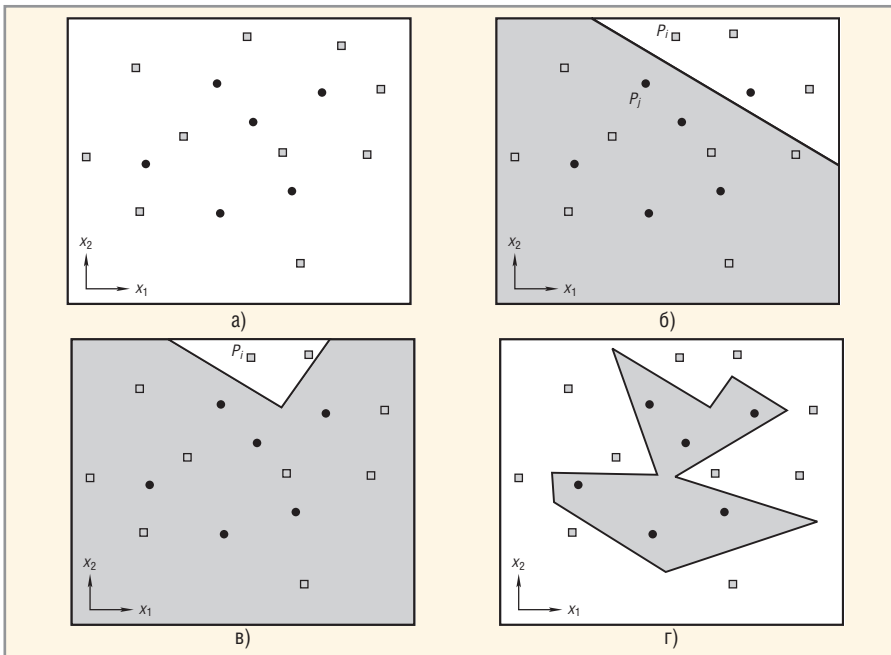


Рис. 5. Построение нейронной сети

(а) Пример обучающей выборки для двумерного входного пространства. Здесь чёрными кружками обозначены точки с $y^* = +1$, серыми квадратами обозначены точки с $y^* = -1$.

(б) Построение разделяющей поверхности между точкой p_i ($y^* = -1$) и точкой p_j ($y^* = +1$). Серым цветом показана аппроксимированная область.

(в) Построение оптимальной разделяющей поверхности для точки p_i ($y^* = -1$). Серым цветом показана аппроксимированная область.

(г) Оптимальная разделяющая поверхность, реализуемая нейронной сетью.

(т.е. перпендикулярной отрезку, соединяющему точки i и j).

Алгоритм: Для любой точки t , принадлежащей искомой гиперплоскости, справедливо следующее утверждение:

$$d_{it} = d_{jt}$$

Отсюда получим:

$$d_{it}^2 = d_{jt}^2,$$

$$\sum_{k=1}^N (x_k^i - x_k^t)^2 = \sum_{k=1}^N (x_k^j - x_k^t)^2,$$

$$\sum_{k=1}^N \left[x_k^t (2x_k^j - 2x_k^i) + \left((x_k^i)^2 - (x_k^j)^2 \right) \right] = 0,$$

$$\sum_{k=1}^N \left(x_k^t (x_k^j - x_k^i) \right) + \frac{1}{2} \sum_{k=1}^N \left((x_k^i)^2 - (x_k^j)^2 \right) = 0. \quad (2)$$

Так как x_k^t – независимая переменная, то (2) можно переписать в виде (1), где:

$$w_0 = \frac{1}{2} \sum_{k=1}^N \left((x_k^i)^2 - (x_k^j)^2 \right),$$

$$w_k = x_k^j - x_k^i, k = 1..N \quad (3)$$

2. Реализация инвариантности к яркости изображения

Дано: $(x_1^i, \dots, x_N^i), (x_1^j, \dots, x_N^j)$, – координаты точек i и j в N -мерном прост-

ранстве. Точка i принадлежит классу A , точка $j \in B$.

Требуется: найти коэффициенты w_k ($k = 0..N$) гиперплоскости (1), разделяющей точки i и j , с реализацией инвариантности к яркости изображений i и j .

Алгоритм: Если точка (x_1^i, \dots, x_N^i) принадлежит классу A , то все точки $(x_1^i + q, \dots, x_N^i + q)$ также принадлежат классу A для любых вещественных значений q . Это утверждение справедливо и для точки $j \in B$. То есть из инвариантности к яркости следует, что каждая точка i в многомерном входном пространстве определяет прямую (на рис. 4б это пунктир), и если $j \in A$, то любая точка i , лежащая на этой прямой, также принадлежит классу A . Значит, и уравнение оптимальной разделяющей гиперплоскости при реализации инвариантности к яркости и без неё будет выглядеть по-разному (см. рис. 4в–4д).

Рассмотрим методику построения оптимальной разделяющей гиперплоскости с учётом инвариантности к яркости изображения. Если мы найдём координаты $(x_1^i, \dots, x_N^i) = (x_1^i + q, \dots, x_N^i + q)$ точки i (см. рис. 4б), находящейся на минимальном расстоянии от точки j , то, воспользовавшись ме-

тодической из п. 1, мы получим искомую разделяющую гиперплоскость. По сути нам необходимо найти q .

Искомая гиперплоскость, а точнее, её коэффициенты w_k ($k = 0..N$), должны удовлетворять следующей системе линейных уравнений:

$$\begin{cases} w_0 + w_1 x_1 + w_2 x_2 + \dots + w_N x_N = 0 \\ w_0 + w_1 (x_1 + q) + w_2 (x_2 + q) + \dots + w_N (x_N + q) = 0 \end{cases}$$

Вычтя из второго уравнения первое, получим:

$$\sum_{k=1}^N w_k = 0. \quad (4)$$

Из (3) следует, что:

$$w_k = x_k^j - (x_k^i + q), k = 1..N.$$

Отсюда и из (4) получим:

$$\sum_{k=1}^N x_k^j - x_k^i - q = 0,$$

$$\left[\sum_{k=1}^N (x_k^j - x_k^i) \right] - Nq = 0,$$

$$q = \frac{\sum_{k=1}^N (x_k^j - x_k^i)}{N}. \quad (5)$$

Итак, уравнение оптимальной разделяющей гиперплоскости при инвариантности к яркости изображения имеет вид:

$$w_0 = \frac{1}{2} \sum_{k=1}^N \left((x_k^i + q)^2 - (x_k^j)^2 \right),$$

$$w_k = x_k^j - (x_k^i + q), k = 1..N. \quad (6)$$

3. Построение НС

В качестве функции активации для нейронов сети будем использовать следующую функцию:

$$y(g) = \begin{cases} -1, & g < 0 \\ +1, & g \geq 0 \end{cases}$$

где $g(x) = w_0 + \sum_{k=1}^N w_k x_k$.

Пусть у чёрных точек на рисунке 5а желаемый выход будет $y^* = +1$, а у серых точек желаемый выход будет $y^* = -1$.

Рассмотрим алгоритм построения НС:

- 1) Возьмём из обучающей выборки серую точку p_i ($y^* = -1$). Найдём несколько ближайших к ней чёрных точек p_j ($y^* = +1$). Между серой точкой p_i и каждой чёрной точкой p_j проводим разделяющую гиперплоскость и подсчитываем количе-

ство правильно аппроксимированных точек i -го класса (серых квадратиков). Выбираем ту гиперплоскость, которой соответствует наибольшее количество аппроксимированных точек i -го класса (рис. 5б). Опять находим несколько неправильно аппроксимированных точек j -го класса, ближайших к точке p_i . Между точкой p_i и каждой выбранной точкой p_j проводим разделяющую гиперплоскость и подсчитываем количество правильно аппроксимированных точек i -го класса. Выбираем ту гиперплоскость, которой соответствует наибольшее количество аппроксимированных точек i -го класса. И так действуем до тех пор, пока в области точки p_i не останется неправильно аппроксимированных точек p_j . Очевидно, все эти разделяющие гиперплоскости вокруг точки p_i реализуют выпуклую область, и, объединив выход соответствующих нейронов простой логической операцией ИЛИ, мы получим двухслойную НС, реализующую разделяющую поверхность на рис. 5в. В первом слое НС весами

нейронов будут коэффициенты разделяющих гиперплоскостей, а во втором слое – все веса от нейронов первого слоя будут равны 1, а w_0 будет равен $K - 0,5$, где K – количество нейронов первого слоя.

2) Возьмём следующую точку p_i ($y^* = -1$), попавшую в серую область, и аналогичным образом построим для неё двухслойную НС, реализующую выпуклую область во входном пространстве. Эти две двухслойные НС объединяем в одну неполносвязную трёхслойную НС, добавив единичные веса от нейронов второго слоя к одному нейрону третьего слоя, а смещение третьего слоя настраиваем на реализацию логической операции И. И так далее, пока не получим внутри аппроксимированной (серой) области только чёрные точки (рис. 5г).

Структура полученной нейронной сети показана на рисунке 6. Таким образом, мы имеем трёхслойную неполносвязную НС, в которой каждый вход связан со всеми нейронами первого слоя, каждая группа нейронов первого слоя соединена с одним ней-

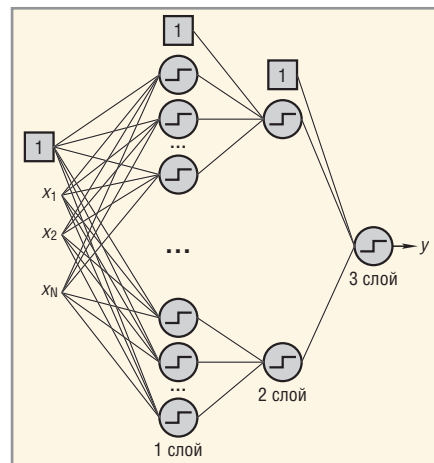


Рис. 6. Структура нейронной сети

роном второго слоя, и все нейроны второго слоя соединены с единственным нейроном третьего слоя.

Далее мы рассмотрим два примера решения задач обработки изображений в нейросетевом логическом базисе: задачу распознавания туберкулезных бактерий на цифровом изображении, полученном с микроскопа, и задачу выделения лиц людей на изображении. ©

Продолжение следует

SHARP

Изображение японского качества

Цветные жидкокристаллические дисплеи с размерами диагонали от **3,5** до **12,1** дюйма для промышленных применений

- Разрешение: 240x320, 320x240, 400x240, 640x480, 800x600 пикселей
- Яркость: до 500 кд/м²

- Широкий угол обзора 140° (по горизонтали)/110° (по вертикали)
- Сменная система задней подсветки на основе люминесцентных ламп с холодным катодом и светодиодов

- Диапазон рабочих температур от -30 до +85°С
- Устойчивость к вибрационным и ударным воздействиям

Дистрибьютор ЖК-дисплеев фирмы Sharp – компания ПРОСОФТ
(495) 234-0636 • info@prosoft.ru • www.prosoft.ru

PROSOFT®