

ПОЭТАПНОЕ ОБУЧЕНИЕ РАДИАЛЬНЫХ НЕЙРОННЫХ СЕТЕЙ

Козин Н.Е., Фурсов В.А.

Институт систем обработки изображений РАН
Самарский государственный аэрокосмический университет

Аннотация

Рассматривается двухэтапная процедура обучения радиальных нейронных сетей распознаванию образов. На первом этапе множество обучающих примеров разбивается на подмножества, для которых настраиваются центры классов. Затем для каждого подмножества обучающих примеров настраиваются параметры радиальных функций. Для формирования подмножеств используются показатели сопряженности векторов признаков обучающих примеров.

1. Постановка задачи

При построении классификаторов широко используются радиальные нейронные сети. Применяемые в них радиальные базисные функции, обычно, имеют локальный характер, что облегчает выбор начальных условий для процесса обучения, обеспечивающего быстрое достижение оптимального решения. Вместе с тем, указанное обстоятельство является причиной более слабой обобщающей способности радиальных сетей по сравнению, например, с сигмоидальными.

Один из путей устранения этого недостатка - разбиение классов обучающих примеров на подклассы. При этом каждый имеет свой «центр», вокруг которого реализуется радиальная базисная функция. Аргументом базисной функции которой является евклидово расстояние образца \mathbf{x} от центра [1]:

$$\psi(\mathbf{x}) = \psi\left(\|\mathbf{x} - \mathbf{c}_{i,j}\|, \sigma_{i,j}\right) \quad (1)$$

Здесь $\mathbf{c}_{i,j}$ - вектор, являющийся образом центра для i -го подкласса j -го класса, $\sigma_{i,j}$ - параметр определяющий допустимую степень отличия предъявляемых образцов от эталона $\mathbf{c}_{i,j}$.

Указанный способ повышения качества распознавания связан с увеличением настраиваемых параметров (центров подклассов). При этом, с одной стороны, существенно возрастают вычислительные затраты при обучении, с другой стороны, падает точность достижения оптимальных решений вследствие увеличения размерности задачи.

В настоящей работе исследуется в значительной степени свободная от указанных недостатков двухэтапная процедура обучения. На первом этапе формируются подклассы и осуществляется «настройка» центров $\mathbf{c}_{i,j}$ для всех подклассов каждого класса. Затем осуществляется обучение нейронных сетей с целью настройки параметров $\sigma_{i,j}$ отдельно для каждого подкласса.

Эффективность указанной схемы в значительной степени связана с критерием, определяющим правило формирования подклассов. Если в качестве критерия использовать, например, число верных и/или ошибочных классификаций, то потребуются многократное повторение указанных этапов настройки

центров $\mathbf{c}_{i,j}$ и параметров $\sigma_{i,j}$. Поэтому наибольший интерес представляют критерии, в которых используются характеристики близости обучающих примеров внутри подклассов.

Эти критерии связаны с качеством распознавания, поскольку более компактное расположение образов в подклассе позволяет затем более точно настроить параметры ($\sigma_{i,j}$). Указанная связь служит основанием для формирования подклассов и настройки их центров $\mathbf{c}_{i,j}$ один раз на начальном этапе. Это позволяет значительно сократить время обучения, т.к. вся технология реализуется за «один проход», а число обучающих образцов в каждом из подклассов оказывается сравнительно небольшим вследствие разбиения классов.

В настоящей работе исследуются процедуры формирования подклассов, основанные на анализе взаимной сопряженности векторов, являющихся образами обучающих примеров.

2. Критерии включения обучающих примеров в подклассы

Пусть \mathbf{X} - матрица, векторы-столбцы которой являются образами обучающих объектов из некоторого класса. Задача заключается в том, чтобы разбить эту матрицу на блоки, каждый из которых включает наиболее близкие в некотором смысле векторы. Эти блоки и будут являться искомыми подклассами обучающих объектов.

В общем случае число векторов в подклассе может быть различным. Задача одновременного выбора оптимального числа и размеров подклассов является сложной. В настоящей работе мы рассмотрим более простую задачу разбиения класса обучающих примеров на заданное число подклассов одинаковых размеров.

Пусть размерность матрицы \mathbf{X} - $N \times M$, где N - число компонент вектора признаков, а M - число векторов-столбцов, которые требуется разбить на k одинаковых подматриц размерности $N \times M_l$, $M_l = M/k$. Для простоты полагаем, что M_l - целое число. Другими словами, необходимо построить систему множеств $\{K_1, \dots, K_l, \dots, K_k\}$, являющуюся разбиением множества \mathbb{N} векторов-столбцов матрицы \mathbf{X} .

Рассматривается три типа критериев, основанных на различных показателях взаимной сопряженности векторов признаков [2]:

1. Показатель максимальной средней парной сопряженности.

При использовании этого показателя ищется такое разбиение множества \mathfrak{N} , при котором сумма коэффициентов парной сопряженности векторов по всем возможным способам разбиения максимальна:

$$\overline{P(k)} = \max \sum_{l=1}^k \rho_l, \quad (2)$$

где

$$\rho_l = \frac{1}{C_{M_l}^2} \sum_{i,j} |r_{i,j}(l)|, \quad l = \overline{1, k},$$

$$r_{i,j}(l) = \frac{[\mathbf{x}_i^T(l) \mathbf{x}_j(l)]}{[\mathbf{x}_i^T(l) \mathbf{x}_j(l)]^2 [\mathbf{x}_i^T(l) \mathbf{x}_j(l)]^2}.$$

Здесь $\mathbf{x}_i(l)$, $\mathbf{x}_j(l)$ - i -ый и j -ый векторы из множества K_l , а $C_{M_l}^2$ - общее число коэффициентов парной сопряженности векторов в множестве K_l (число сочетаний из M_l по 2).

Эта мера является привлекательной с точки зрения вычислительной простоты. Однако, при использовании меры близости (2) из подкласса могут быть исключены векторы, которые вносят незначительный вклад в величину средней парной сопряженности, хотя этот вектор и остальные векторы подкласса являются почти компланарными. От этого недостатка свободен следующий критерий.

2. Показатель максимальной средней сопряженности с подпространством.

Этот показатель определяется как максимальное суммарное значение показателей сопряженности всех векторов подкласса с подпространствами, натянутыми на векторы остающиеся в подклассе после исключения этого вектора при всех возможных способах разбиения:

$$\overline{R(k)} = \max \sum_{l=1}^k R_l, \quad (3)$$

$$R_l = \frac{1}{M_l} \sum_i |R_i(l)|, \quad l = \overline{1, k},$$

$$R_i = \mathbf{x}_i^T(l) \mathbf{X}_{M_{i,j-1}}(l) \left[\mathbf{X}_{M_{i,j-1}}^T(l) \mathbf{X}_{M_{i,j-1}}(l) \right]^{-1} \mathbf{X}_{M_{i,j-1}}^T(l) \mathbf{x}_i(l)$$

где $\mathbf{x}_i(l)$ - i -ый вектор из множества K_l , а $\mathbf{X}_{M_{i,j-1}}(l)$ - $N \times (M_{i,j-1})$ - матрица, составленная из остальных векторов-столбцов из указанного множества K_l .

Геометрически этот показатель косинус угла между i -м вектором (\mathbf{x}_i) и подпространством, натянутым на векторы-столбцы, оставшиеся в подматрице размерности $N \times (M_l - 1)$, после исключения из нее этого i -го вектора.

3. Показатель минимальной средней сопряженности с нуль-пространством.

Показатель вычисляется как минимальное суммарное значение показателей сопряженности всех векторов подкласса с нуль-пространствами подпространств, натянутых на векторы, остающиеся в подклассе после исключения вектора, для которого вычисляется показатель:

$$\overline{S(k)} = \min \sum_{l=1}^k S_l, \quad (4)$$

$$S = \frac{1}{M_l} \sum_i |S_i(l)|, \quad l = \overline{1, k},$$

$$S_i = \frac{[\mathbf{x}_i^T \mathbf{T}_{M_{j-1}} \mathbf{T}_{M_{j-1}}^T \mathbf{x}_i]^{\frac{1}{2}}}{[\mathbf{x}_i^T \mathbf{x}_i]^{\frac{1}{2}}},$$

где $\mathbf{T}_{M_{j-1}}$ - $N \times (N - M_{j-1} + 1)$ -матрица, составленная из $N - M_{j-1} + 1$ собственных векторов, соответствующих нулевым собственным значениям матрицы $\mathbf{X}_{M_{j-1}} \mathbf{X}_{M_{j-1}}^T$, полученной из исходной матрицы \mathbf{X}_{M_j} вычеркиванием вектора-столбца \mathbf{x}_i . Геометрически S_i - косинус угла между вектором \mathbf{x}_i и нуль-пространством подпространства, натянутого на остальные векторы матрицы \mathbf{X}_{M_j} [3].

По существу показатели R и S являются мерами одной и той же характеристики, но вычисляются с использованием разных подпространств. Ясно, что показатель минимальной средней сопряженности с нуль-пространством выгоднее использовать в тех случаях, когда размерность нуль-пространств меньше, чем размерности столбцовых пространств матриц \mathbf{X}_{M_j} .

3. Итерационная процедура формирования подклассов

Для вычисления приведенных выше мер близости векторов в подклассах, строго говоря, необходимо для каждого класса перебирать все возможные сочетания по k векторов в подклассах. Для этого потребуются огромный объем вычислений. Предлагается следующая простая итерационная схема формирования подклассов.

На первом шаге с использованием одного из указанных выше критериев или путем непосредственного анализа изображений определяются векторы, являющиеся по предположению центрами подклассов. В частности, в могут быть взяты два наиболее «удаленных» по критерию парной сопряженности вектора x_i и x_j , для которых значение r взаимной сопряженности минимально:

$$r = \min_{i,j} \frac{x_i^T x_j}{\sqrt{(x_i^T x_j)} \sqrt{(x_j^T x_i)}}.$$

Указанные векторы используются в качестве начальных центров «притяжения» подклассов. На следующем шаге среди оставшихся $M-2$ векторов определяется по одному ближайшему вектору к каждому из найденных на первом шаге, по одному из указанных выше критериев. Далее шаги повторяются до тех пор, пока либо каждый элемент исходного множества не будет соотнесён с одним из подклассов, либо в каждом подклассе не окажется заданное количество векторов. В первом случае исходное множество будет разделено на конечное число подмножеств, не пересекающихся между собой.

Следует заметить, что размерность подпространства, натянутого на векторы, входящие в подклассы, растёт от шага к шагу. Поэтому с вычислительной точки зрения начиная с некоторого шага может оказаться более выгодным вместо среднего значения показателей минимальной сопряженности (2) использовать среднее показателей максимальной сопряженности с нуль-пространством (3).

После того как подклассы сформированы, для каждого из них определяется вектор c , являющийся характеристикой «центра» подкласса. В простейшем случае он может быть вычислен как среднее арифметическое всех элементов подкласса:

$$c_j = \frac{1}{N} \sum_{i=1}^N x_{ij},$$

где N – число элементов данного подкласса. Для формирования центров подклассом могут использоваться алгоритмы, построенные на основе принципа согласованности [3].

4. Общая схема обучения радиальной нейронной сети

Общая схема радиальной нейронной сети с поэтапным обучением показана на рис. 1. Предполагается, что для каждого распознаваемого объекта формируется несколько векторов, характеризующих центры. Входной вектор одновременно подаётся на все нейроны входного слоя с центрами, отвечающими соответствующим подклассам.

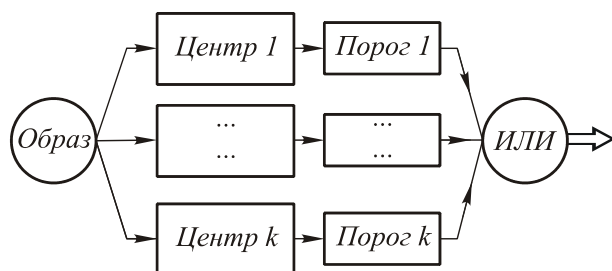


Рис. 1. Общая структура нейронной сети

После сравнения входного вектора с каждым центром, величины отклонений от центров сравниваются с допустимыми значениями, определяемыми пороговыми элементами. Значения порогов в общем случае различны для каждого центра. По результатам сравнения принимается решение по схеме ИЛИ (“OR”). Если хотя бы в одном случае отклонение от

центра не превышает порога, принимается решение об отнесении входного образа к данному классу. В противном случае формируется сигнал, соответствующий отказу от распознавания.

Геометрически (рис. 2) область, соответствующая элементам данного класса, является объединением областей, соответствующих сформированным подклассам.

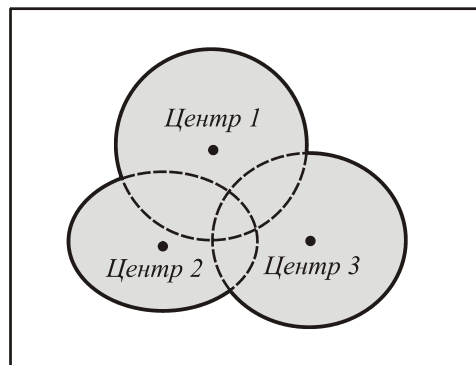


Рис. 2. Пример объединения областей вокруг центров

Использование нейронных сетей с радиальными базисными функциями описанной архитектуры наиболее эффективно в случае, когда предъявляются жесткие требования к вероятности ложных классификаций. В данном случае надежность классификатора по этому показателю может быть достигнута путем увеличения числа подклассов и изменения порогов.

5. Пример

В качестве примера ниже приводятся результаты экспериментов связанных с задачей распознавания стилизованной цифры «9». Для сравнения проводилось два эксперимента: для случая одного центра каждого класса и для случая двух центров каждого класса.

Пример полученных центров для цифры «9» представлен на рис.3.

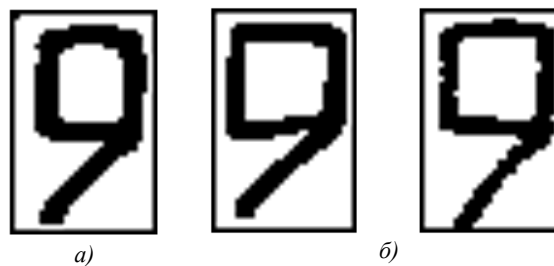


Рис. 3. – Изображения центров для случаев а) одного центра б) двух центров

Для оценивания эффективности необходимым являлось условие, что обучающая и тестовая выборки должны были быть одними и теми же в случаях и первого и второго этапов. Тестирование сети проводилось на выборке, состоящей из выбранных случайным образом 360 образцов. Обучающая выборка состояла из 36 образцов также выбранных произвольно.

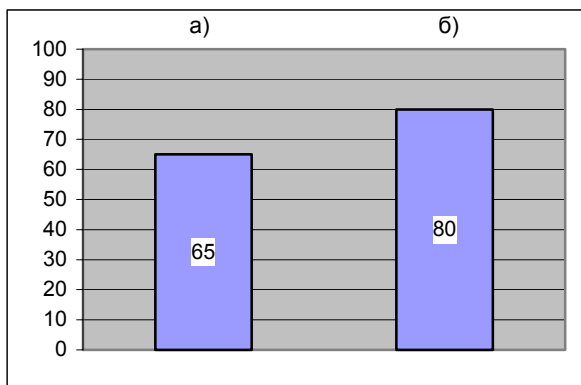


Рис. 4. Процент правильных распознаваний для случая:
а) одного центра; б) двух центров

Процент правильных распознаваний, полученный на контрольной выборке, для случаев одного и двух центров приведен на рис. 4.

7. Заключение

Приведенные результаты показывают возможность увеличения качества распознавания с использованием описанной методики.

Ясно, что для достижения высокой эффективности потребуется существенное увеличение числа подклассов для каждого класса (типа объекта). Исследования показывают, что при добавлении новых центров к уже существующим, необходимо использовать обучающие примеры, для которых число отказов в распознавании велико.

8. Благодарности

Работа выполнена при поддержке Министерства образования РФ, Администрации Самарской области, Американского фонда гражданских исследований и развития (CRDF) и РФФИ (гранты №01-01-00097 №03-01-00109).

Литература

1. Осовский С. Нейронные сети для обработки информации // Пер. с польского И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с.
2. Методы компьютерной обработки изображений // Под ред. В.А. Сойфера. – М.: Физматлит, 2001. – 784 с.
3. Fursov, Vladimir A. Constructing unified identification algorithms using a small number of observations for adaptive control and navigation systems // Journal: Proc. SPIE Vol. 3087, P. 34-44.