Вычислительные технологии **Том 6, № 1, 2001**

ВЫБОР МИНИМАЛЬНОЙ КОНФИГУРАЦИИ НЕЙРОННЫХ СЕТЕЙ

Н. А. Игнатьев
Национальный университет Узбекистана, Ташкент
e-mail: tin000@tashsu.silk.org

A method for the selection of the neuron networks configuration is suggested for solving the problems of pattern recognition implemented through the combination of the selection processes of the set of informative signs kits and the selection of the minimum coverage of the training set by standards.

Введение

Одним из общих свойств всех нейронных сетей (НС) является параллельная обработка сигналов, для реализации которой необходимо разбиение множества нейронов на слои и соединение определенным образом нейронов различных слоев, а в ряде случаев и нейронов одного слоя между собой. Обоснование необходимых и достаточных свойств сети для решения того или иного рода задач — важнейший этап разработки нейрокомпьютерной техники.

Отсутствие фундаментальных теоретических разработок не позволяет производить синтез НС в строгой зависимости от решаемой задачи. В большинстве случаев параметры нейронных сетей с фиксированной структурой (конфигурацией) настраиваются для решения конкретной задачи и оптимальный вариант получается на основе интуитивного подбора [1]. Поиск минимального числа нейронов и связей между ними, достаточных для решения задачи, определяет процесс выбора минимальной конфигурации НС. Этот процесс не противоречит идее бритвы Оккама, смысл которой заключается в том, что при отсутствии каких-либо специальных указаний из множества возможных решений предпочтение отдается простейшему из них.

В настоящей статье поиск минимальной конфигурации НС для решения задач распознавания образов предлагается проводить посредством отбора информативных наборов признаков и соответствующего ему минимального покрытия эталонами объектов обучающей выборки. Для направленного отбора информативных признаков вычисляются их веса с заданными ограничениями [2]. Эти веса в дальнейшем используются для определения величин синаптических связей нейронов. Решение задачи минимального покрытия объектов обучающей выборки эталонами базируется на идее метода линейных оболочек [3].

Под линейной оболочкой понимается множество граничных по заданной мере близости (метрике) объектов классифицированных выборок. Линейные оболочки получили свое название по аналогии с кусочно-линейным решающим правилом, которое геометрически интерпретирует использование евклидовой метрики для классификации по минимуму расстояния. В [3] показано, что с помощью линейных оболочек можно получить локально-оптимальное покрытие обучающей выборки эталонами. Распознавание по этим эталонам с помощью правила “ближайший сосед” является корректным на объектах обучения. Естественный компромисс между числом эталонов (кусков гиперплоскостей) и размерностью признакового пространства может быть разрешен через вычисление меры статистического разнообразия распознающих алгоритмов [4].

1. Компактизация классифицированных выборок объектов

Различные методы и алгоритмы теории распознавания образов базируются на гипотезе о компактности. К их числу относится метод вычисления весов признаков, детально изложенный в [2]. Согласно этому методу, на множестве объектов обучения $E_0 = \{S_1, ..., S_m\}$ с описаниями в R^n, разделенном на l непересекающихся классов $K_1, ..., K_l$, вводится функционал

$$J(w) = \frac{\sum\limits_{i=1}^{n} w_i\theta_i}{\sum\limits_{i=1}^{n} w_i\gamma_i} \to \min, \tag{1}$$

в котором θ_i, γ_i есть меры соответственно внутриклассового сходства и межклассового различия по i-му признаку. Значения весов признаков w_i, $i = \overline{1,n}$ интерпретируются как коэффициенты сжатия (растяжения) координатных осей признакового пространства и вычисляются построением функции Лагранжа

$$F(w, \lambda) = \frac{\sum\limits_{i=1}^{n} w_i\theta_i}{\sum\limits_{i=1}^{n} w_i\gamma_i} + \lambda\left(\sum_{i=1}^{n} w_i - 1\right)$$

на функционале (1) при заданных ограничениях $\sum\limits_{i=1}^{n} w_i = 1$, $w_i \geq 0$:

$$w_i = \begin{cases} \dfrac{\gamma_i - \theta_i}{\sum\limits_{\{j|\gamma_j-\theta_j>0\}} \gamma_j - \theta_j}, & \gamma_i - \theta_i \ > 0, \\ 0, \quad \gamma_i - \theta_i \ \leq 0. \end{cases} \tag{2}$$

Выделение множества объектов линейных оболочек (граничных объектов) классов проводится с целью вычисления весов и отбора информативных наборов признаков, а также построения минимального покрытия обучающей выборки эталонами.

Пусть в качестве меры близости между объектами выбрана взвешенная евклидова метрика

$$\rho(x, y) = \sqrt{\sum_{i=1}^{n} w_i^2 (x_i - y_i)^2}, \tag{3}$$

в которой значения весов вычислены по формуле (2). Для каждого объекта $S_i \in K_j \cap E_0$ построим последовательность объектов E_0, упорядоченных, согласно метрике (3), по возрастанию расстояния от объекта S_i:

$$S_{i_0}, S_{i_1}, ..., S_{i_{m-1}},$$

где $S_{i_0} = S_i$.

Пусть $S_{i_c} \in K_u$, $u \neq j$, $c \in \{1, ..., m-1\}$ — ближайший к S_i объект, не входящий в класс K_j. Обозначим через $O(S_i)$ окрестность радиуса $\rho(S_i, S_{i_c})$ с центром в S_i, включающую все объекты, для которых $\rho(S_i, S_{i_t}) < \rho(S_i, S_{i_c})$, $t = \overline{1, c-1}$. Из $O(S_i)$ найдем такой объект S_{i_r}, $r \in \{0, ..., c-1\}$, что

$$\rho(S_{i_c}, S_{i_r}) = \min_{S_{i_t} \in O(S_i)} \rho(S_{i_c}, S_{i_t}). \tag{4}$$

Множество объектов $L(E_0, l) = \{S_{i_r}\}$, определяемых на E_0 по (4), назовем линейной оболочкой классов K_1, ..., K_l. Из различных способов вычисления значений θ_i, γ_i выберем два:

а)

$$\theta_i = \sum_{u=1}^{l} \sum_{S_j \in K_u} | a_{ij} - a'_{ui} |,$$

$$\gamma_i = \frac{1}{l-1} \sum_{u=1}^{l} \sum_{S_j \notin K_u} | a_{ij} - a'_{ui} |,$$

где $S_j = (a_{j1}, ..., a_{jn})$, $a'_{ui} = \sum\limits_{S_t \in K_u} a_{ti} / \mid K_u \mid$;

б)

$$\theta_i = \sum_{u=1}^{l} \sum_{S_j \in K_u} | a_{ji} - b_{ji} |,$$

$$\gamma_i = \sum_{u=1}^{l} \sum_{S_j \notin K_u} | a_{ji} - c_{ji} |,$$

где $(b_{j1}, ..., b_{jn})$ и $(c_{j1}, ..., c_{jn})$ — значения признаков ближайших к $S_j \in K_q$, $q = \overline{1, l}$ объектов соответственно множеств $L(E_0, l) \cap K_q$ и E_0 / K_q.

Обозначим через $H(r) = (x_1, ..., x_r)$, $2 \leq r \leq n$ набор из r признаков и примем, что w_i есть значение веса признака $x_i \in H(r)$. Процесс определения значений $\{w_i\}$ по (2) на различных наборах $H(r)$ существенно сокращается при выборе способа *а*), так как значения $\{\theta_i\}$, $\{\gamma_i\}$ нужно вычислять только один раз. Естественное условие эффективности этого процесса — несовпадение математических ожиданий признаков классов. Количество вычислений $\{\theta_i\}$, $\{\gamma_i\}$ значительно возрастает при способе *б*), поскольку для каждого набора признаков $H(r)$ их нужно выполнять заново. Значение разности $\gamma_i - \theta_i \neq \mathrm{const}$, так как число и состав объектов линейных оболочек на различных наборах признаков не являются фиксированными.

В [2] показано, что процесс вычисления значений весов $\{w_i\}$, $i = \overline{1, r}$ можно совмещать с направленным отбором информативных наборов признаков. При определенных ограничениях на веса признаков из набора $H(r)$ всегда можно удалить такой признак, что значение функционала (1) на $H(r-1)$ будет меньше, чем на $H(r)$. Положим, что значения

$\{\theta_i\}$, $\{\gamma_i\}$ вычислены по способу a), веса признаков $\{\omega_i\}$, определенных на наборе $H(r)$, $2 \le r \le n$, — по формуле (2) и $w_i > 0$, $\sum\limits_{x_i \in H(r)} w_i = 1$ для всех $x_i \in H(r)$. Тогда имеет место

Теорема. *Необходимым и достаточным условием выбора признака $x_j \in H(r)$ в качестве кандидата на удаление из набора $H(r)$ является*

$$\frac{\theta_j}{\gamma_j} - \frac{\sum\limits_{x_i \in H(r)} w_i \theta_i}{\sum\limits_{x_i \in H(r)} w_i \gamma_i} = \max_{H(r)}.$$

Доказательство. Обозначим

$$\frac{c_1}{c_2} = \frac{\sum\limits_{x_i \in H(r)} w_i \theta_i}{\sum\limits_{x_i \in H(r)} w_i \gamma_i}.$$

С учетом (2) это отношение запишем

$$\frac{c_1}{c_2} = \frac{\sum\limits_{x_i \in H(r)} \theta_i(\gamma_i - \theta_i)}{\sum\limits_{x_i \in H(r)} \gamma_i(\gamma_i - \theta_i)}.$$

Поскольку признак $x_j \in H(r)$ — кандидат на удаление, то

$$\frac{c_1}{c_2} - \frac{c_1 - \theta_j(\gamma_j - \theta_j)}{c_2 - \gamma_j(\gamma_j - \theta_j)} > 0.$$

В силу заданных на $\{w_i\}$ ограничений $\forall x_i \in H(r) \omega_i > 0$, $\sum\limits_{x_i \in H(r)} \omega_i = 1$ получим

$$\frac{\theta_j}{\gamma_j} > \frac{c_1}{c_2},$$

а с учетом (1)

$$\frac{\theta_j}{\gamma_j} - \frac{\sum\limits_{x_i \in H(r)} w_i \theta_i}{\sum\limits_{x_i \in H(r)} w_i \gamma_i} = \max_{H(r)}.$$

Теорема доказана.

2. О минимальном покрытии обучающей выборки эталонами

Формально классификация объектов с помощью нейронных сетей происходит путем разбиения гиперпространства, определяемого информативным набором признаков $H(r)$, рядом гиперплоскостей, число которых зависит от решения задачи минимального покрытия эталонами обучающей выборки. Поиск этого числа параллельно с вычислением весов и отбором информативных наборов признаков может проводиться в форме решения задачи минимального покрытия обучающей выборки эталонами.

Весовой вектор линейного дискриминанта может быть вычислен с помощью взвешенного евклидова расстояния (3) между объектом $S = (a_1, ..., a_n)$ и эталоном $B = (b_1, ..., b_n)$ [5]:

$$\varphi(S, B) = \sum_{i=1}^{n} v_i a_i + v_0, \quad v_i = w_i^2 b_i, \quad v_0 = -\frac{1}{2} \sum_{i=1}^{n} w_i^2 b_i^2.$$

Рассмотрим конечно-сходящийся алгоритм [3] для локально-оптимального покрытия эталонами обучающей выборки. Определение множества объектов линейной оболочки $L(w, l, r) = \{Y_1^1, ..., Y_d^1\}$, $d \leq m$ с помощью весов (2) на наборе признаков $H(r)$, $2 \leq r \leq n$ считаем первым шагом этого алгоритма. Для каждого $Y_i^1 \in L(w, l, r) \cap K_j$, $j = \overline{1, l}$ упорядочим объекты класса K_j по возрастанию расстояния (3) от Y_i^1:

$$S_{i_1}, ..., S_{i_{m_j}}, \tag{5}$$

где $S_{i_1} = Y_i^1$, $m_j = \mid K_j \mid$. На t-м шаге ($t = 2, 3, ...$) для очередного последовательно предъявляемого эталона Y_i^{t-1} вычислим

$$Y_i^t = (r_i^{t-1} Y_i^{t-1} + S_{i_{r_i^{t-1}}})/(r_i^{t-1} + 1), \tag{6}$$

где r_i^{t-1} — число объектов из упорядоченного набора (5), используемых для вычисления Y_i^{t-1}, $r_i^1 = 1$.

Если построенный по эталонам $Y_1^{t-1}, ..., Y_i^t, ..., Y_d^{t-1}$ распознающий алгоритм остается корректным (не делающим ошибок) на E_0, то новое значение i-го эталона находим по формуле (6), $r_i^t = r_i^{t-1} + 1$, $Y_\mu^t = Y_\mu^{t-1}$, $r_\mu^t = r_\mu^{t-1}$, $\forall \mu \neq i$.

Проверка условия о том, что на очередном t-м шаге ни для одного из эталонов $Y_1^{t-1}, ..., Y_d^{t-1}$ применение (6) не дает корректного алгоритма на E_0, служит критерием останова вычислительного процесса. Сходимость алгоритма к локально-оптимальному покрытию обучающей выборки эталонами за конечное число шагов доказана в [3].

Пусть $\{Y_1, ..., Y_u\}, u \leq d$ — минимальное покрытие эталонами выборки E_0 по $H(r)$, $\omega = (\omega_1, ..., \omega_n)$ — информационный вектор, задаваемый на наборе признаков $\{x_1, ..., x_n\}$, при этом $\omega_i = 1$, если i-й признак входит в набор $H(r)$, и $\omega_i = 0$, если не входит в него, ωS, ωY_j представляют ω — части объекта S и эталона Y_j. Максимальное значение линейного дискриминанта $\varphi(\omega S, \omega Y_j)$, вычисляемое по $\{Y_1, ..., Y_u\}$, определяет номер эталона j, к классу которого относят объект S.

Очевидно, что количество эталонов, получаемое с помощью различных способов вычисления значений $\{\theta_i\}$, $\{\gamma_i\}$ на различных информативных наборах признаков, не фиксировано. Выбор оптимального (допустимого) решения основывается на вычислении емкости кусочно-линейных решающих функций [4] и определяется из условия

$$kr \to \min_{E_0},$$

где k — число эталонов (кусков гиперплоскостей) на обучающей выборке; r — размерность пространства информативного набора признаков.

При использовании метода линейных оболочек традиционный подсчет количества ошибок на обучающей и контрольной (тестовой) выборках для проверки качества итоговой модели алгоритмов распознавания приобретает новый смысл. Поскольку распознающий алгоритм, определяемый на эталонах $\{Y_1, ..., Y_u\}$, корректен на обучающей выборке по построению, то большое количество ошибок на контрольной выборке указывает на то, что объекты обучения плохо представляют (через линейную оболочку) конфигурацию классов. Так как процесс формирования выборок объектов имеет случайный характер, то необходима проверка гипотезы о том, что при росте объема E_0 межклассовое расстояние удовлетворяет соотношению

$$\frac{1}{|L(\omega, l, r)|} \sum_{j=1}^{l} \sum_{S_i \in L(\omega,l,r)\cap K_j} \min_{S_q \in E_0 \setminus K_j} \rho(S_i, S_q) \to \text{const}.$$

Если объекты классов линейно не разделимы в исходном признаковом пространстве, то существует возможность их разделения путем перехода в обобщенное пространство. На практике для таких целей чаще всего используются квадратичные решающие правила, задаваемые поверхностями второго порядка. Переход с линейной решающей функции на квадратичную связан с большими затратами вычислительных ресурсов. Объем вычислений можно значительно сократить, если использовать такую эвристическую процедуру.

Пусть v — максимальное число обобщенных признаков, с которого начинается поиск оптимальной конфигурации нейронной сети. Определим максимальное число r ($r \le n$), удовлетворяющее неравенству

$$\frac{r(r+3)}{2} \le v,$$

с помощью которого найдем информативный набор признаков $H(r) = (x_1, ..., x_r)$. Используя различные комбинации признаков из $H(r)$ в степени не выше 2 по описанной выше схеме, можно получить оптимальную конфигурацию нейронной сети.

Список литературы

[1] Горбань А. Н., Россиев Д. А. Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996.

[2] Игнатьев Н. А. Компактизация классифицированных выборок объектов // Методы и вычислительные средства обработки видеоинформации, данных и анализа. Ташкент, 1993. С. 75–82.

[3] Игнатьев Н. А. Некоторые вопросы реализации комбинированных систем распознавания в классе линейных решающих функций // Вопр. вычисл. и прикл. математики. Ташкент, 1995. Т. 100. С. 88–94.

[4] Вапник В. Н. Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979.

[5] Ту Дж., Гонсалес Р. Принципы распознавания образов. М.: Мир, 1978.

Поступила в редакцию 12 апреля 2000 г.
в переработанном виде — 31 июля 2000 г.