V. N. Vapnik, A. Ya. Chervonenkis, A class of algorithms for pattern recognition learning, *Avtomat. i Telemekh.*, 1964, Volume 25, Issue 6, 937–945

Том XXV «АВТОМАТИКА И ТЕЛЕМЕХАНИКА» № 6

1964

УДК 62.506.1 + 621.391.193

ОБ ОДНОМ КЛАССЕ АЛГОРИТМОВ ОБУЧЕНИЯ РАСПОЗНАВАНИЮ ОБРАЗОВ

В. Н. ВАПНИК, А. Я. ЧЕРВОНЕНКИС

(*Москва*)

Формулируются две задачи опознания: нахождение образа обучающейся системой и синтез обучающейся системы. Для первой задачи приводится формальная схема обучения и получены оценки длины обучающей последовательности. Предлагается эвристическая программа для решения задачи синтеза обучающейся системы.

1. Задача обучения распознаванию образов интуитивно ставится так: образ мыслится как некоторое множество элементов, а обучение узнаванию как умение находить все множество после предъявления его части.

Такая постановка задачи с формальной точки зрения не может считаться удовлетворительной, пока не наложено ограничение на класс возможных множеств объектов, определяемых как образ. Само введение ограничений является не формальным и должно быть основано на некоторой не формальной гипотезе. Такая гипотеза содержится во всех алгоритмах распознавания, хотя может быть и в неявном виде.

После того как ограничение введено, обучение распознаванию можно описать следующей схемой. Имеется устройство, или автомат, распознающий некоторую систему образов, т. е. разбивающий множество входных ситуаций на классы. Обучающееся устройство наблюдает в течение некоторого времени последовательность входных ситуаций, возникающих на входе этого автомата, и его реакции. Этот период времени называется периодом обучения, а последовательность входных ситуаций и реакций — обучающей последовательностью. По окончании периода обучения обучающееся устройство переходит в некоторое состояние, в котором оно осуществляет определенную свою классификацию входных ситуаций. Задача обучения состоит в том, чтобы перевести обучающееся устройство в такое состояние, в котором оно будет осуществлять разбиение, достаточно близкое в определенном смысле к разбиению, осуществляемому исходным автоматом. Если это условие выполнено, то говорят, что устройство обучилось имитации исходного автомата или же распознаванию системы образов, задаваемой имитируемым автоматом.

При создании обучающихся устройств приходится исходить из двух противоречивых требований. С одной стороны, требуется, чтобы оно могло обучаться имитации достаточно широкого класса автоматов, а с другой стороны, чтобы обучение происходило за не слишком большое число показов. Если обучающееся устройство способно обучиться имитации некоторого класса автоматов, то каждому из этих автоматов (или группе близких автоматов) должно соответствовать определенное состояние обучающегося устройства, и обучение сводится к выбору этого состояния из всех возможных на основании обучающей последовательности. Пусть число возможных состояний равно N, тогда для выбора одного из них, вообще говоря, необходимо сообщить $\lg N$ *бит* информации. В частности, если входная ситуация описывается бинарным вектором размерности n и имитируемые автоматы осуществляют дихотомию множества входных ситуаций, то в принципе возможно 2^{2n} разбиений. Если не наложить более никаких ограничений, то потребовалось бы около 2^n показов, что неприемлемо. Здесь приходится применить те неформальные ограничения, о которых говорилось выше.

Представляется целесообразным разделить задачу распознавания на две.

Первая задача — синтез обучающегося устройства — состоит в первоначальном отборе системы допустимых разбиений, соответствующей набору состояний обучающегося устройства. Она решается конструктором на основании неформальных ограничений. Попытка решить эту задачу на основании обучающей последовательности приводит, очевидно, к требованию перебора всех возможных входных ситуаций.

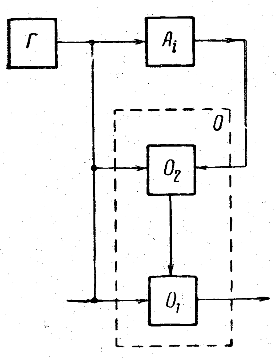

Рис. 1

Вторая задача, решаемая обучающимся устройством, состоит в выборе определенного состояния из заданного набора на основании обучающей последовательности. Это — задача имитации конкретного автомата из заданного класса.

В настоящей статье рассматривается формальная постановка второй задачи, исследуется более подробно такой класс алгоритмов работы обучающегося устройства, при котором по окончании обучения выбирается такое состояние, что обучающееся устройство классифицирует входные ситуации, предъявленные при обучении безошибочно. Показывается, что длина обучающейся последовательности может быть ограничена в этом случае сверху и оценка при достаточно больших N имеет порядок $\lg N$, где N — число возможных состояний. Полученная оценка применяется к персептронам, обучаемым по методу, описанному в работе [2]. Предлагается также некоторая эвристическая программа для синтеза обучающегося устройства.

2. Рассмотрим формальную схему, описывающую обучение автомата распознаванию образов (рис. 1).

Задан набор L автоматов A_i, каждый из которых осуществляет разбиение некоторого множества входов E на два класса, т. е.

$$A_i(x) = 1, \text{ если } x \in M_i, \quad M_i \bigcup M_i' = E;$$

$$A_i(x) = 0, \text{ если } x \in M_i, \quad M_i \bigcap M_i' = 0.$$

Такие автоматы в дальнейшем будем называть распознающими. Каждому автомату A_i поставлено в соответствие вероятностное распределение P_i на множестве входов E, определяющее статистику его входов.

Обучающееся устройство O состоит из двух частей: части O_1, для которой задан набор состояний $\{S_k\}$, в каждом из которых часть O_1 представляет собой распознающий автомат на множестве входов E, и части O_2 — обучателя. Часть O_2 работает лишь в процессе обучения, при этом обучатель наблюдает работу одного из автоматов класса L, т. е. на вход автомата A_i подается некоторая последовательность $x_1, \ldots, x_m$, $x_i \in E$ и эта же последовательность вместе с ответами автомата A_i идет на вход обучателя O_2. Таким образом, вход обучателя есть последовательность

$$(x_1, y_1), \quad (x_2, y_2), \ldots, \quad (x_m, y_m),$$

где $x_i \in E$, а y_i — соответствующие выходы автомата A_i. Выходами обучателя являются сигналы S_k, переводящие O_1 в состояние S_k набора $\{S_k\}$.

Обучение распознаванию образа, задаваемого автоматом (имитации автомата A_1) на последовательности $x_1, \ldots, x_m$, будем считать удачным, если вероятность различных ответов автомата A_i и $S_k = O_2(x_1, A_i(x_1), \ldots$ $\ldots, x_m, A_i(x_m))$ будет меньше наперед заданной величины $\varkappa$ при подаче на вход этих автоматов случайного элемента x при распределении P, генерируемого блоком Г.

Пусть

$$A_i(x) = 1, \quad \text{если} \quad x \in M_i; \; S_k(x) = 1, \quad \text{если} \quad x \in M_{S_k};$$
$$A_i(x) = 0, \quad \text{если} \quad x \in M_i'; \; S_k(x) = 0, \quad \text{если} \quad x \in M'_{S_k},$$

тогда вероятность различных ответов A_i и S_k

$$P_{A_i S_k} = P(M_i \cap M'_{S_k}) + P(M_i' \cap M_{S_k})$$

и обучение удачно, если $P_{A_i S_k} < \varkappa$.

Способность обучающегося устройства O обучиться имитации любого автомата из класса L с точностью $\varkappa$ на обучающей последовательности длины m можно было бы определить следующим образом.

1. Для всякого автомата $A_i \in L$ существует такая последовательность $x_1, \ldots, x_m$ элементов из E, что часть O_1 в состоянии $S_k = O_2(x_1, A_i(x_1), \ldots$ $\ldots, x_m, A_i(x_m))$ дает различные ответы с автоматом A_i на случайный элемент x с вероятностью $P_{A_i S_k} < \varkappa$.

Для каждого m при заданной $\varkappa$ можно определить подмножество L_m класса L такое, что обучающееся устройство способно обучиться имитации любого автомата из L_m на последовательности показов длины не больше m. Индекс m определяет величину естественности подкласса L_m для обучающегося устройства O.

Приведенное выше определение оставляет открытым вопрос о том, как найти обучающую последовательность. Поэтому можно предложить другое, более узкое определение способности обучающегося устройства O обучиться имитации любого автомата из класса L.

2. Для всякого автомата $A_i \in L$ и для любой случайной последовательности $x_1, \ldots, x_m$ длины m независимых выборок при распределении P с вероятностью, большей $1 - \eta$, $P_{A_i S_k}$ будет меньше $\varkappa$, где

$$S_k = O_2(x_1, A_i(x_1), \ldots, x_m, A_i(x_m)).$$

Т. е., предполагая, что обучение производится на любой случайной последовательности, обладающей той же статистикой, что и последовательность предъявлений на экзамене, требуем, чтобы на последовательности длины m с вероятностью, большей $1 - \eta$, устройство O обучилось имитации любого автомата $A_i \in L$ с точностью $\varkappa$ $(P_{A_i S_k} < \varkappa)$.

Рассмотрим теперь случай, когда множества возможных состояний обучающегося устройства и множество имитируемых автоматов совпадают и состоят из N элементов

$$L = \{S_k\} = (A_1, \ldots, A_N).$$

Пусть алгоритм обучателя O_2 таков, что по предъявленным объектам $x_1, \ldots, x_m$ и ответам $y_1, \ldots, y_m$ имитируемого автомата A_i он выбирает такое состояние устройства O_1, что

$$A_i(x_k) = A_1(x_k) = y_k \text{ для всех } k(1 \leqslant k \leqslant m).$$

Подобный алгоритм назовем алгоритмом с полной памятью.

Теорема 1. Обучающееся устройство, реализующее алгоритм с полной памятью, способно имитировать любой автомат из $A_1, \ldots, A_N$ в смысле определения (2) с точностью $\varkappa$ с вероятностью, большей $1 - \eta$, при длине обучающей последовательности

$$m = \frac{\ln N - \ln \eta}{\ln (1 - \varkappa)}.$$

Действительно, пусть автомат A_j, выбранный обучателем O_2, таков, что $P_{A_iA_j} > \varkappa$. Тогда вероятность того, что на случайной последовательности $x_1, \ldots, x_m$ они дадут одинаковые ответы

$$P_{ij}^{(m)} = (1 - P_{A_iA_j})^m \leqslant (1 - \varkappa)^m.$$

Вероятность того, что хотя бы один автомат A_j такой, что $P_{A_iA_j} \geqslant \varkappa$, дал одинаковый ответ с A_i на случайной последовательности длины m будет

$$P_i^{(m)} \leqslant \sum_{A_j (P_{A_iA_j} \geqslant \varkappa)} (1 - P_{A_iA_j})^m < N (1 - \varkappa)^m.$$

Потребуем, чтобы $P^{(m)} < \eta$. Для этого достаточно

$$N(1 - \varkappa)^m = \eta \quad \text{или} \quad m = (\ln N - \ln \eta) / -\ln (1 - \varkappa).$$

Для малых $\varkappa$ получаем

$$m = (\ln N - \ln \eta) / \varkappa.$$

Рассмотренная теорема оставляет открытым вопрос о том, как построить алгоритм с полной памятью, не содержащий полного перебора состояний из L. В ряде случаев, однако, такие алгоритмы удается построить [1, 2].

Интересно отметить, что приведенная оценка не может быть существенно улучшена без привлечения дополнительных сведений относительно плотности распределения и множества L.

Рассмотрим следующий частный случай. Пусть из N автоматов $A_1, \ldots, A_N$ набора L вероятность выдачи одинакового ответа автоматов $A_1, \ldots, A_{N-1}$ с имитируемым равна $(1 - \varkappa)$ и вероятность выдачи одинакового ответа автомата A_N с имитируемым равна единице. Определим, какой длины должна быть обучающая последовательность, для того чтобы с достоверностью, равной $1 - \eta$, утверждать, что обучатель выбрал автомат A_N.

Для этого воспользуемся формулой вероятности гипотез Бейеса, где будем полагать априори вероятность выбора автомата $P(A_i) = P(A_j)$. В этом случае вероятность того, что один из автоматов $A_1, \ldots, A_{N-1}$ выдаст на обучающей последовательности длины $\overline{m}$ те же ответы, что и имитируемый, равна

$$P(\overline{m}) = \frac{(N-1)(1-\varkappa)^{\overline{m}}}{(N-1)(1-\varkappa)^{\overline{m}} + 1}.$$

Потребуем, чтобы $P(\overline{m}) < \eta$.

Для этого достаточно, чтобы

$$\frac{(N-1)(1-\varkappa)^{\overline{m}}}{(N-1)(1-\varkappa)^{\overline{m}} + 1} < \eta,$$

откуда

$$\overline{m} > \left[\ln \frac{\eta}{1-\eta} - \ln (N-1)\right] \Big/ \ln (1 - \varkappa).$$

Сравним оценку, полученную в этом частном случае, с оценкой, получающейся из общих предположений:

$$\frac{\overline{m}}{m} - 1 = \left[\ln (1-\eta) - \ln\left(1 - \frac{1}{N}\right)\right] \Big/ (\ln \eta - \ln N).$$

Из сравнения видно, что при η, близких к нулю, и достаточно больших N эти две оценки практически совпадают.

Однако ясно, что существуют такие плотности распределения, для которых оценка длины обучающей последовательности может быть существенно улучшена. Но поскольку априори плотности распределения вероятности нам, как правило, неизвестны, доуточнение оценки может проходить по мере обучения автомата.

Рассмотрим обучающееся устройство, которое одновременно с обучением проводит проверки. Проверка состоит в сравнении выходов имитируемого автомата и автомата A_j, выбранного к началу проверки, на последовательности длины $m(k)$, где k — номер проверки. Если в течение проверки выходы сравниваемых автоматов совпадают, то устройство заканчивает обучение, в противном случае обучение продолжается.

Справедлива следующая теорема.

Теорема 2. Если обучающееся устройство заканчивает обучение, то с вероятностью, большей $1-\eta$, выбранный автомат имитирует заданный с точностью $\varkappa$ при условии, что

$$m(k)=\frac{\ln\eta-\ln\zeta(n)-n\ln k}{\ln(1-\varkappa)},$$

где n — любая константа $(n>1)$; $\zeta(n)=\sum_{i=1}^{\infty}\frac{1}{i^n}$.

В самом деле, предположим, что проверяли автоматы $A_1,\ldots,A_j$. Пусть P_i есть вероятность различных ответов автомата A_i и имитируемого.

Предположим, что для всех автоматов $P_i>\varkappa$.

Тогда вероятность P того, что хотя бы один из автоматов $A_1,\ldots,A_k,\ldots$ даст те же ответы, что и имитируемый, равна

$$P=\sum_k(1-P_k)^{m(k)}\leqslant\sum_k(1-\varkappa)^{m(k)}.$$

Пусть функция $m(k)$ такова, что $(1-\varkappa)^{m(k)}=\alpha/k^n$ или $m(k)=(\ln\alpha-n\ln k)/\ln(1-\varkappa)$.

Определим α из условия $P\leqslant\eta$:

$$P\leqslant\sum_k(1-\varkappa)^{m(k)}\leqslant\sum_k\frac{\alpha}{k^n}=\alpha\zeta(n)=\eta.$$

Отсюда $\ln\alpha=\ln\eta-\ln\zeta(n)$, где $\zeta(n)$ — дзета-функция Римана. При этом

$$m(k)\geqslant\frac{\ln\eta-\ln\zeta(n)-n\ln k}{\ln(1-\varkappa)}.$$

Оценка справедлива для любого n.

При $n=2$

$$m(k)\geqslant\frac{\ln\eta-\ln\frac{\pi^2}{2}-2\ln k}{\ln(1-\varkappa)}.$$

Для случая, когда обучатель меняет автомат только при совершении им ошибок, число k равно числу ошибок с начала обучения.

Рассмотрим множество автоматов, дающее одинаковые ответы с имитируемым на некоторой обучающей последовательности длины m.

Пусть $L_{m-1}\neq L_m$. Очевидно, $L_m\subset L_{m-1}$. Рассмотрим последовательность вложенных множеств

$$L_{m_1}\supset L_{m_2}\supset,\ldots,\supset L_{m_k}.$$

Теорема 2 позволяет при заданном η поставить в соответствие множеству L_{m_i} вероятность $(1-\varkappa_i)$.

3. Разбиение обучающегося устройства на части O_1 и O_2 является чисто формальным, и в конкретных реализациях они могут быть совмещены.

Рассмотрим обучение персептрона с точки зрения этой формальной схемы. Каждое состояние персептрона как распознающей машины при заданной коммутации определяется набором весов $(\lambda_1, \ldots, \lambda_n)$, с которыми суммируются возбуждения нейтронов в R-клетке. Будем считать, что часть O_2 представляет персептрон целиком, вместе с блоком выработки весов λ_i, а часть O_1 является копией персептрона O_2 без блока выработки весов. Набором состояний $\{S_k\}$ является множество векторов $(\lambda_1, \ldots, \lambda_n)$, сигналом S_k — выработанная обучателем система весов λ. Имитируемым автоматом при обучении персептрона обычно бывал человек, множеством входов — набор картинок на некотором рецепторном поле. Реакциями человека на обучающую последовательность, предъявляемую персептрону, служат сигналы поощрения.

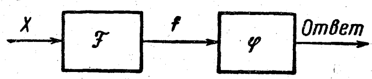

Рис. 2

Однако персептрон [3] не реализует алгоритм с полной памятью, так как обучающая последовательность после обучения по статистической схеме не обязательно опознается с достоверностью, равной единице. Метод обучения, реализующий алгоритм с полной памятью, описан в [1, 2], где рассматривается схема (рис. 2), на которой X — функция, заданная на рецепторном поле, а f — функция после стандартного преобразования (отображения над рецепторным полем).

В [2] в качестве функции f выбран вектор размерности n, координаты которого принимают значения 0 или 1. В качестве автоматов $A_1, \ldots, A_N$ набора L выбраны сечения гиперкуба n — 1-мерной гиперплоскостью. Там же предложен метод построения гиперплоскости, основанный на нахождении обобщенного портрета образа.

Выбирать в качестве автоматов сечения гиперплоскостью n-мерного куба удобно, потому что, во-первых, построение таких сечений относительно несложно, и, во-вторых, количество их у n-мерного куба относительно невелико и любое сечение может быть найдено согласно приведенной оценке при помощи реального числа показов.

В самом деле, пусть множество автоматов $A_1, \ldots, A_N$ есть множество пороговых элементов с произвольными весами $\lambda_1, \ldots, \lambda_n$ и порогом θ, реализующее все возможные логические функции n переменных, различные между собой (и реализуемые сечения n-мерного куба n — 1-мерной гиперплоскостью).

Число таких пороговых элементов равно числу компонентов, на которые разбивают $n+1$-мерное пространство $\lambda_1, \ldots, \lambda_n$ θ гиперплоскости вида $\Sigma\lambda_i y_i = 0$ по всем 2^n наборам

$$Y=(y_1, \ldots, y_n), \ y_i = \begin{cases} 0, \\ 1, \end{cases} \quad (i=1, \ldots, n).$$

Справедливо следующее рекуррентное соотношение:

$$e_m^{\ r} \leqslant e_m^{r-1} + e_{m-1}^{r-1},$$

где $e_m^{\ r}$ — максимальное число компонентов, на которые r гиперплоскостей размерности $m-1$ разбивают m-мерное пространство $e_1^{\ r} = r+1$, $e_m^{\ 1} = 2$.

Из рекуррентного соотношения видно, что

$$e_m^{\ r} \leqslant e_m^{\ 1} + \sum_{i=2}^{r-1} e_{m-1}^{i},$$

но так как $e_m^1 = e^1_{m-1}$, то

$$e_m^r \leqslant \sum_{i=1}^{r-1} e_{m-1}^i .$$

С другой стороны, при $r \gg m$

$$e_m^r \leqslant \sum_{i=3}^{r-1} e_{m-1}^i \approx \int_0^r e_{m-1}^{(x)} \, dx \approx \frac{r^m}{m!} .$$

В нашем случае $m = n + 1$, $r = 2^n$, откуда

$$e_m^r \leqslant \frac{2^{n(n+1)}}{(n+1)!} .$$

Следовательно, если обучающееся устройство таково, что по заданным входам, каждый из которых есть набор n переменных, и ответу заданного порогового элемента A_j оно выбирает пороговый элемент A_i, дающий те же ответы на эти входы, то в силу общей оценки при заданных $\varkappa$ и η для успешного обучения достаточно иметь обучающую последовательность длины

$$m = \frac{\ln(\eta/N)}{\ln(1-\varkappa)} = \frac{\ln\eta - n(n+1) + \sum_{k=1}^{n} \ln k}{\ln(1-\varkappa)} \approx \frac{\ln\eta - n(n+1)}{\ln(1-\varkappa)} .$$

Из приведенного рассмотрения ясно, что чем больше размерность вектора f, тем большей должна быть длина обучающей последовательности, необходимая для того, чтобы с заданной степенью надежности выбирать автомат, близкий к имитируемому. Однако при любой размерности вектора f алгоритм, рассмотренный в [2], сходится.

4. Рассмотрим автомат Г, который генерирует векторы X, и автомат A, осуществляющий дихотомию этих векторов.

Пусть при помощи автоматов A и Г составлена обучающая последовательность M_m векторов X (индекс m указывает на ее длину).

Автоматами A и Г в задачах опознания обычно являлись человек или группа людей. Решением первой задачи опознания — синтеза автомата — явилось бы нахождение такого алгоритма, который, используя обучающую последовательность M_m $(m \leqslant 2^n)$, из множества всех возможных автоматов отобрал бы некоторое подмножество T (состоящее из t элементов) такое, что с вероятностью P оно содержало бы автомат A (или близкий к нему).

Ясно, что такого алгоритма не существует.

В связи с этим в работах по опознаванию вводились гипотезы, направленные на выявление множества T. Они обычно накладывали ограничения на свойства множеств в рецепторном поле. Другой путь связан с наложением ограничений на процесс построения множества автоматов T. Он приводит к необходимости имитации эволюции.

В основу ограничений положены следующие соображения:

1) изменение структуры автомата случайно;

2) сохранение измененного детерминировано и является функцией приспосабливаемости автомата к среде;

3) характер оценки приспосабливаемости автомата к среде не зависит от автомата, а зависит от среды.

Изложенные соображения до некоторой степени моделируют процесс эволюции.

Согласно работам [1, 2], дихотомия образов $°\Phi_1$ и $°\Phi_2$ может быть произведена тогда и только тогда, когда отображение $\mathcal{F}$ (рис. 2) таково, что векторы $f_i^1 - f_j^2$ лежат внутри выпуклого конуса, где вектор f_k^i поставлен отображением $\mathcal{F}$ в соответствие k-му объекту i-го образа.

В самом деле, если выполняется неравенство $(f^i\varphi_i) > (f^j\varphi_i)$, то $(f^i - f^j\varphi_i) > 0$.

Наоборот, если векторы $f^i - f^j$ лежат внутри выпуклого конуса, то существует опорная плоскость к конусу с нормалью ψ и справедливо $(f^i + f^j, \varphi) > 0$ или $(f^i\psi) > (f^j\psi)$.

Нашей целью является построение такого отображения $\mathcal{F}$, в котором векторы $f^i - f^j$ лежат внутри выпуклого конуса.

Из двух отображений $\mathcal{F}_1$ и $\mathcal{F}_2$ будем считать лучшим то, в котором минимальный раствор конуса, заключающий векторы $f^i - f^j$, меньше. Отображение считается найденным, если минимальный раствор конуса, заключающий векторы, меньше $\pi / 2$.

В качестве оценки приспосабливаемости примем величину раствора наименьшего заключающего конуса. Далее будем полагать, что отображение $\mathcal{F}$ строится из однородных элементов. В частности, ими могут быть пороговые элементы — нейроны.

Под нейронами понимается следующая пороговая модель.

Нейрон имеет n входов $x_1, \ldots, x_n$ и один выход y_j, причем входы и выход связаны соотношением

$$y_j = \theta(\Sigma\xi_i x_i - \xi_0),$$

где коэффициенты ξ могут принимать любые значения $-a \leqslant \xi \leqslant b$, ξ_0 — величина, называемая порогом нейрона. Будем оценивать приспосабливаемость нейрона к среде по величине наименьшего раствора конуса, заключающего векторы $f^i - f^j$, заданные на множестве его входов. Изменение множества входов будем называть изменением нейрона.

В этих терминах процесс построения множества автоматов может быть описан следующим образом.

Пусть заданы обучающая последовательность (среда) длины m и исходное рецепторное поле. Рассматривается образование k нейронов:

1) выходы каждого из k нейронов соединяются со случайно выбранным множеством рецепторов;

2) нейроны адаптируются к среде, т. е. определяются $\xi_1, \xi_2, \ldots, \xi_n, \xi_0$ для каждого нейрона так, чтобы заключить первые m_1 векторов $f^i - f^j$ в выпуклый конус.

Для этого строится вектор $\psi = (\xi_1, \ldots, \xi_n, \xi_0)$ по тому же алгоритму, по которому вырабатывается обобщенный портрет [1, 2].

Вектор ψ может быть построен только на части обучающей последовательности: может возникнуть ситуация, когда вектор существует для показов $x_1, \ldots, x_m$ и не существует для показов $x_1, \ldots, x_m, x_{m+1}$.

В этом случае нейрону присваиваются коэффициенты вектора ψ, образованного на показах $X_1, \ldots, X_{m_1}$

Далее рассматривается расширенное рецепторное поле, состоящее из старого рецепторного поля и выходов образованных нейронов.

Если на расширенном рецепторном поле векторы $f^i - f^j$ не могут быть заключены внутрь выпуклого конуса, то на расширенном рецепторном поле образуются новые нейроны и рецепторное поле вновь расширяется, и т. д.

Пусть после k-го расширения векторы $f^i - f^j$ могут быть заключены внутрь выпуклого конуса. Пусть размерность этих векторов равна d, т. е. найдено пространство H, в котором векторы могут быть разделены.

Интересно найти подпространство пространства H минимальной размерности, в котором эти векторы также могут быть разделены.

К сожалению, вопрос о нахождении такого подпространства остается открытым. Однако можно найти такое подпространство минимальной размерности, в котором проекция минимального заключающего конуса остается выпуклой.

В самом деле, пусть Q — подпространство пространства H, в котором проекция заключающего конуса остается выпуклой (рис. 3). Тогда $\alpha + \beta < \pi / 2$. Но при нормированных векторах и нормированном обобщенном портрете угол $\alpha = \arccos C_i$, где C_i — порог обобщенного портрета.

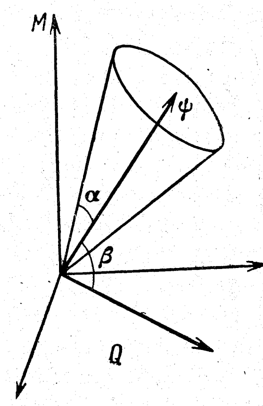

Рис. 3

Следовательно, условием, при котором конус остается выпуклым, является условие

$$\beta < \frac{\pi}{2} - \arccos C_i \text{ или}$$

$$\cos\beta > \cos\left(\frac{\pi}{2} - \arccos C_i\right).$$

Это значит, что проекция обобщенного портрета на подпространство должна быть больше некоторой заданной величины.

Алгоритм выбора подпространства следующий. На пространстве размерности d строится обобщенный портрет $\psi = (\xi_1, \ldots$ $\ldots, \xi_d, \xi_0)$. Определяется величина $\gamma = 1 - \cos\beta = 1 - \cos(\pi / 2 -$ $- \arccos \xi_0)$. Отбираются наименьшие значения ξ_i^2 координаты так, чтобы $\sqrt{\xi_n^2 + \ldots + \xi_{n_i}^2} \leqslant \gamma$. Пусть эти координаты образуют подпространство M. Найденным подпространством является подпространство $Q =$ $= H / M$.

Цитированная литература

1. Вапник В. Н., Лернер А. Я. Узнавание образов при помощи обобщенных портретов. Автоматика и телемеханика, т. XXIV, № 6, 1963.
2. Вапник В. Н., Червоненкис А. Я. Об одном классе персептронов. Автоматика и телемеханика, т. XXV, № 1, 1964.
3. Розенблат Ф. Обобщение восприятий по группам преобразований. Кибернетический сборник, № 4, 1962.

A CLASS OF ALGORITHMS FOR PATTERN RECOGNITION LEARNING

V. N. VAPNIK, A. Ya. CHERVONENKIS

Two recognition problems are formulated: the finding of a pattern by a learning system, the synthesis of a learning system. A formal scheme of a learning is described for the first problem. Estimates of learning sequence duration are obtained. An euristic programm for the solution of problem of learning system synthesis is proposed.