ՀԱՅԿԱԿԱՆ ՍՍՀ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԿԱԴԵՄԻԱ ՀԱՅԱՍՏԱՆԻ ԿԵՆՍԱԲԱՆԱԿԱՆ ՀԱՆԴԵՍ
АКАДЕМИЯ НАУК АРМЯНСКОЙ ССР. БИОЛОГИЧЕСКИЙ ЖУРНАЛ АРМЕНИИ

XIX, № 11, 1966

А. В. МЕВИС

ОБ ОДНОМ МЕТОДЕ ПОСТРОЕНИЯ АДАПТИВНОЙ СИСТЕМЫ

Адаптивные системы — это новый класс систем переработки информации, которые по своим структурно-функциональным свойствам и способностям организации обработки информации приближаются к биологическим системам, в частности, центральной нервной системе человека и высших позвоночных животных. Будучи приближенными моделями биологических систем, адаптивные автоматы воспроизводят незначительную часть функциональных свойств и многообразных форм поведения биологических прототипов. Но, тем не менее, построение адаптивных систем в настоящее время представляет большой интерес как с теоретической, так и с практической точек зрения.

При построении адаптивных систем возникают две проблемы: какая должна быть классифицирующая функция и как должен проходить процесс обучения. В статье приводится математическая модель адаптивной системы, состоящая из блоков распознавания и обучения. Обсуждается метод построения классифицирующей функции, так как такая функция наиболее легко реализуется с помощью простых пороговых схем. Приводятся основные положения, каким должна удовлетворять обучающаяся схема.

Математическая модель адаптивной системы. Пусть адаптивная система состоит из схем распознавания и схемы обучения, как показано на рис. 1. На вход схемы распознавания подается вектор $x_j = (x_1, x_2, \cdots, x_n)$, $j = 1, 2, \cdots, 2^n$, составляющие которого прини-

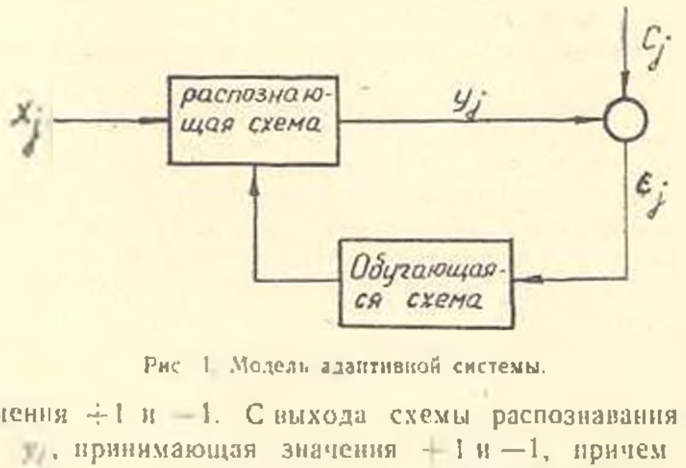

Рис. 1. Модель адаптивной системы.

мают значения +1 и −1. С выхода схемы распознавания снимается величина y_j, принимающая значения +1 и −1, причем $y_j = f(x_j)$, $j = 1, 2, \cdots, 2^n$. Величина y_j поступает на схему сравнения с желаемым выходным сигналом $c_j = z(x_j)$, $j = 1, 2, \cdots, 2^n$, поступающим иавие. Выходной сигнал со схемы сравнения поступает на вход схемы обучения, выходной сигнал которой по цепи обратной связи изменяет параметры схемы распознавания так, чтобы величины y_i и c_j совпадали для максимального числа входных сигналов.

Схема распознавания реализует некоторую функцию $y_i = f(x_i)$, называемую классифицирующей функцией. Для того чтобы описать поведение схемы распознавания, ниже приводится синтез классифицирующей функции.

Синтез классифицирующей функции. Рассмотрим статистический метод Chow [2]. Пространство входных сигналов состоит из всех вершин n-мерного куба, где любая вершина соответствует некоторому сигналу. Процесс классификации множества сигналов $\{x_i\}$, принимающих два значения 0 или 1, или -1 и $+1$, сводится к разбиению пространства сигналов на два непересекающихся множества.

Пусть $p(x_i/c_j)$ — условная вероятность того, что множество сигналов $\{x_i\}$ относится к классу c_j, $j = 1, 2$. $p_j = p(\{x_i\})$ — априорная вероятность того, что множество сигналов $\{x_i\}$ относится к классу c_j, $j = 1, 2$. Очевидно, что $p_1 + p_2 = 1$, где p_1 — априорная вероятность того, что множество $\{x_i\}$ принадлежит к классу c_1; p_2 — априорная вероятность того, что множество $\{x_i\}$ принадлежит к классу c_2.

Chow вводит алгоритм распознавания, обеспечивающий минимум вероятности ошибки и состоящий в отнесении множества $\{x_i\}$ к классу c_1, если

$$p_1 p(\{x_i\}/c_1) > p_2 p(\{x_i\}/c_2) \tag{1}$$

и в отнесении к классу c_2, если

$$p_1 p(\{x_i\}/c_1) < p_2 p(\{x_i\}/c_2). \tag{2}$$

Причем критерий минимума вероятности ошибки при разбиении пространства входных сигналов на два непересекающихся множества V_1 и V_2 определяется из условия:

$$V_1 = \left\{ \frac{\{x_i\}}{p_1 p(\{x_i\}/c_1) > p_2 p(\{x_i\}/c_2)} \right\}, \tag{3}$$

$$V_2 = \left\{ \frac{\{x_i\}}{p_1 p(\{x_i\}/c_1) < p_2 p(\{x_i\}/c_2)} \right\}.$$

На основе такого разбиения булева функция $F(x_n)$ n переменных может быть представлена в следующем виде:

$$F(x_1, x_2, \ldots, x_n) = \begin{cases} 1, \text{ если } x \in V_1 \\ 0, \text{ если } x \in V_2 \end{cases} \tag{4}$$

Пусть $F(x_n) = y_j$, т. е. равна классифицирующей функции. В таком случае любая задача двоичного распознавания с помощью указанных распределений определяет единственную булеву функцию [3].

Частный случай двоичного распознавания имеет место при статистической независимости переменных $\{x_{ij}\}$, т. е. когда условные вероятности $p(\{x_i\}/c_1)$ и $p(\{x_i\}/c_2)$ определяются следующим образом:

$$p(\{x_i\}/c_j) = \prod_{i=1}^{n} p(x_i/c_j),\ j=1,\ 2. \tag{5}$$

Покажем, что в этом случае классифицирующая функция есть пороговая функция, т. е. требуется показать, что если имеет место соотношение (5), то функция, определяемая в выражении (4), будет пороговой.

Булева функция $F(x_1, x_2, \cdots, x_n)$ n переменных $x_1, x_2, \cdots, x_n$ является пороговой, если существует совокупность $(n+1)$ действительных чисел $\omega_0, \omega_1, \cdots, \omega_n$, для которых справедливы следующие неравенства:

$$\left.\begin{aligned} &\omega_0 + \sum_{i=1}^{n} \omega_i x_i > 0,\ \text{при}\ F(x_1, x_2, \cdots x_n) = 1 \\ &\omega_0 + \sum_{i=1}^{n} \omega_i x_i < 0,\ \text{при}\ F(x_1, x_2, \cdots x_n) = 0 \end{aligned}\right\} \tag{6}$$

Из выражения (4) булева функция n переменных имеет вид:

$$F(x_1, x_2, \cdots x_n) = \begin{cases} 1,\ \text{если}\ p_1 \prod_{i=1}^{n} p(x_i/c_1) > p_2 \prod_{i=1}^{n} p(x_i/c_2) \\ 0,\ \text{если}\ p_1 \prod_{i=1}^{n} p(x_i/c_1) < p_2 \prod_{i=1}^{n} p(x_i/c_2) \end{cases} \tag{7}$$

Перепишем выражение (7) так:

$$F(x_1, x_2, \cdots x_n) = \begin{cases} 1,\ \text{если}\ \log p_1 + \sum_{i=1}^{n} \log p(x_i/c_1) > \log p_2 + \sum_{i=1}^{n} \log p(x_i/c_2) \\ 0,\ \text{если}\ \log p_1 + \sum_{i=1}^{n} \log p(x_i/c_1) < \log p_2 + \sum_{i=1}^{n} \log p(x_i/c_2) \end{cases} \tag{8}$$

Пусть x_i принимают значения 0 и 1. Тогда

$$p(x_i/c_j) = \begin{cases} t_{ij},\ \text{если}\ x_i = 0, \\ 1 - t_{ij},\ \text{если}\ x_i = 1, \end{cases} \tag{9}$$

или

$$p(x_i/c_j) = t_{ij}^{(1-x_i)} (1 - t_{ij})^{x_i}. \tag{10}$$

Далее

$$\log p(x_i/c_j) = \log t_{ij} + x_i \log \frac{1 - t_{ij}}{t_{ij}}, \tag{11}$$

теперь

$$F(x_1, x_2, \cdots, x_n) = \begin{cases} 1,\ \text{если}\ \omega_0 + \sum_{1}^{n} \omega_i x_i > 0, \\ 0,\ \text{если}\ \omega_0 + \sum_{1}^{n} \omega_i x_i < 0. \end{cases} \tag{12}$$

что и требовалось доказать, где

$$\omega_0 = \log\frac{p_1}{p_2} + \sum_{i=1}^{n}\log\frac{\lambda_{i1}}{\lambda_{i2}}\,, \tag{13}$$

$$\omega_i = \log\frac{(1-\lambda_{i2})\lambda_{i2}}{(1-\lambda_{i1})\lambda_{i1}}\,. \tag{14}$$

Отсюда следует, что для простой технической реализации схем распознавания наиболее приемлемыми являются линейные пороговые классифицирующие функции, так как они требуют только суммирования простых произведений вместо умножения большого числа значений $p(x_i/c_j)$.

Таким образом, адаптивную модель удобно строить на адаптивных пороговых элементах (рис. 2). Выходная величина пороговых

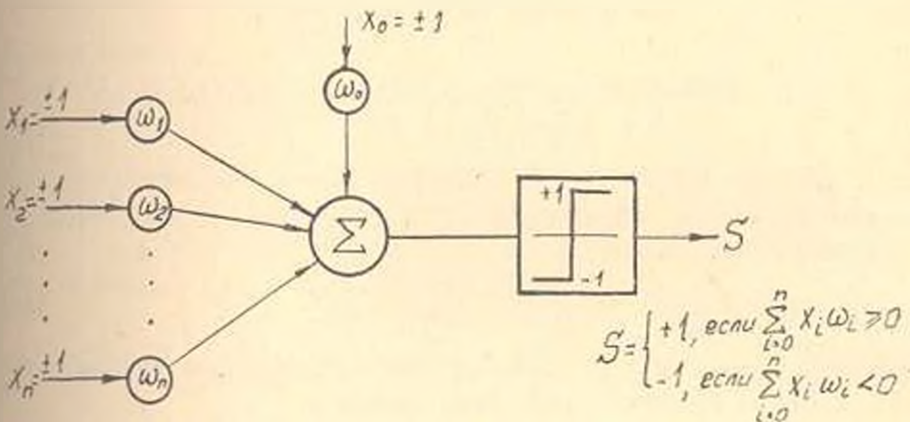

Рис. 2. Адаптивный пороговой элемент.

элементов зависит от образования взвешенной суммы входных величин: если взвешенная сумма больше или равна пороговой величине, то на выходе имеем 1, если меньше пороговой величины, на выходе имеем 0 или —1. Входные переменные x_i умножаются на весовые величины ω_i и результирующие произведения суммируются. Аналоговая сумма S посылается на квантующее устройство, образуя двоичный сигнал.

В работе [4] показано, что целью адаптации порогового элемента при заданном наборе входных сигналов и соответствующих им выходных сигналов является нахождение множества весовых векторов $\omega_0, \omega_1, \cdots, \omega_n$, минимизирующих среднеквадратичную ошибку ε^2 отклонения между выходным и желаемым выходным сигналами. При этом показано, что значение ε^2 является квадратичной функцией настроек весовых величин.

Построение схемы распознавания. На основе вышесказанного, можно представить схему распознавания, как показано на рис. 3. На вход этой схемы подается последовательность двоичных сигналов, характеризующих с помощью следующих вероятностей:

$$p\,(x_j / c_j = +1) = p\,(x_j)\,p\,(c_j = +1 / x_j),$$

$$p\,(x_j / c_j = +1) = p\,(x_j)\,p\,(c_j = -1 / x_j),$$

где $p\,(x_i / c_j)$ — условная вероятность появления x_j при $c_j = \pm 1$.

$p\,(x_j)$ — априорная вероятность появления x_j,

$p\,(c_j / x_j)$ — условная вероятность того, что $c_j = \pm 1$ при условии появления x_j.

Введем функцию, которую назовем желаемой выходной функцией

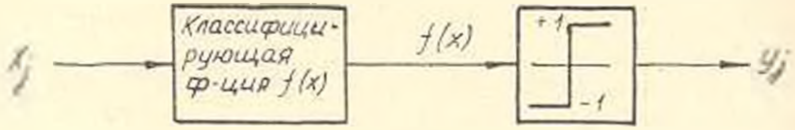

Рис. 3. Модель распознающей схемы.

$$y_j^{\bullet} = p\,(x_j / c_j = +1) - p\,(x_j / c_j = -1) = p\,(x_j)\,[\,p\,(c_j = +1 / x_j) -$$

$$- p\,(c_j = -1 / x_j)] \tag{15}$$

Тогда, согласно (1), (2) и (4), вероятность появления ошибок схемой распознавания будет минимальной, если ее выход определяется из следующего условия:

$$y_j = \begin{cases} +1, & \text{если } y_j^{\bullet} \geqslant 0, \\ -1, & \text{если } y_j^{\bullet} < 0. \end{cases} \tag{16}$$

Ниже приводится таблица значений классифицирующей функции y_j в зависимости от комбинации входных сигналов.

Таблица 1

x	x_1	x_2	— — —	x_n		x_j
$x^{(0)}$	-1	-1	— — —	-1		$y^{\bullet}_{(0)}$
$x^{(1)}$	$+1$	-1	— — —	-1		$y^{\bullet}_{(1)}$
$x^{(2)}$	$+1$	$+1$	— — —	-1		$y^{\bullet}_{(2)}$
.	—	—	— — —	—		.
.	—	—	— — —	—		.
.	—	—	— — —	—		.
$x^{(2^n-1)}$	$+1$	$+1$	— — —	$+1$		$y^{\bullet}_{(2^n-1)}$

Как показано в уравнении (12), классифицирующая функция должна иметь вид:

$$y_j = \sum_{i=0}^{2^n-1} a_i X_i, \tag{17}$$

где X_i образуются комбинацией входных сигналов $x_1, x_2, \cdots x_n$:

$$X_0 = x_0,\ X_{n+1} = x_1 x_2,$$
$$X_1 = x_1,\ X_{n+2} = x_1 x_2 x_3,$$
$$\vdots$$
$$X_n = x_n,\ X_{2^n-1} = x_1 x_2 \cdots x_n.$$

Коэффициенты линейной комбинации ω_l определяются из условия минимизации среднеквадратичного отклонения между y_j и y_j^*:

$$\varepsilon^2 = \sum_{j=0}^{2^n-1} (y_j - y_j^*)^2 = \sum_{j=0}^{2^n+1} \left(\sum_{l \in I} \omega_l X_{l(j)} - y_j^* \right)^2, \tag{18}$$

где

$$I = (i_1, i_2, \cdots, i_m),\ 0 \leqslant l \leqslant 2^n - 1.$$

Как показано в работе [1],

$$\frac{\partial \varepsilon^2}{\partial \omega_k} = -2 X_k \varepsilon(k) = -2 X_k (y_j - y_j^*), \tag{19}$$

где

$$\varepsilon(k) = y_j - y_j^*,\ k = (i_1, i_2, \cdots i_m).$$

Условие $\min \varepsilon^2$ приводит к следующей системе уравнений, если принять во внимание (19):

$$\frac{\partial \varepsilon^2}{\partial \omega_k} = -2 \left[\sum_{l \in I} \omega_l \left(\sum_{j=0}^{2^n-1} X_{l(j)} \cdot X_{k(j)} - \sum_{j=0}^{2^n-1} y_j^* X_{k(j)} \right) \right] = 0, \tag{20}$$

$$k = (i_1, i_2, \cdots i_m).$$

Так как все величины X равны $+1$ или -1, имеет место следующее соотношение:

$$\sum_{j=0}^{2^n-1} X_{l(j)} \cdot X_{k(j)} = \begin{cases} 0, & \text{при } i \neq k, \\ 1, & \text{при } i = k. \end{cases} \tag{21}$$

Следовательно, решением системы уравнений (20) будет

$$\omega_l = \sum_{j=0}^{2^n-1} y_j^* X_{l(j)}. \tag{22}$$

Подставив (15) в (22), получим

$$\omega_l = \sum_{j=0}^{2^n-1} [p(x_j/c_l = +1) - p(x_j/c_l = -1)] X_{l(j)} =$$
$$= \{\sum_{\{X_j=+1\}} [p(x/c = +1) - p(x/c = -1)] - \sum_{\{X_j=-1\}} [p(x/c = 1) - p(x/c = -1)]\} = \tag{23}$$
$$= \{[p(x=1/c=1) - p(x=1/c=-1)] - [p(x=-1/c=1) -$$
$$- p(x=-1/c=-1)]\} = p(x=c) - p(x \neq c).$$

На основе вышесказанного, в литературе [5] приводится выражение для линейной классифицирующей функции, минимизирующей среднеквадратичную ошибку, имеющее следующий вид:

$$y = \sum_{i=0}^{n} w_i x_i = k \sum_{i=0}^{n} [p(x=c) - p(x \neq c)]\, x_i =$$

$$= k \sum_{i=0}^{n} \left(\frac{m^{(+)}}{m} \gamma_i^{(+)} - \frac{m^{(-)}}{m} \gamma_i^{(-)} \right) x_i, \qquad (24)$$

где k — коэффициент пропорциональности,

m — общее число опытов.

$m^{(\pm)}$ — число опытов, для которых $c = \pm 1$,

$\frac{m^{\pm}}{m} = p(c = \pm 1)$,

γ_i — ожидаемое значение x_i при $c = \pm 1$, т. е.

$$\gamma_i = p(x_i = +1/c = \pm 1) - p(x_i = -1/c = \pm 1).$$

Построение схемы обучения. Известно, что на одном пороговом элементе можно реализовать очень ограниченное количество логических функций, более того, для реализации многих функций на одном элементе нужен более длительный процесс адаптации. Схема обучения должна обладать следующими свойствами:

1. Поскольку схема является моделью нервной сети, она должна состоять из большего числа взаимосвязанных однородных элементов, простых и универсальных. Соединение большого количества элементов позволяет получить достаточную гибкость;

2. Желательно схему строить на адаптивных пороговых элементах;

3. Схема должна иметь большое число внутренних параметров, обладающих способностью перестраиваться при наличии сигналов ошибки, что дает возможность схеме приспосабливаться к изменяющимся соотношениям между входным и выходным сигналами. Процесс изменения параметров схемы под действием сигнала ошибки можно рассматривать как процесс адаптации схемы к изменению среды.

Московский авиационный институт,
Лаборатория нейробионики
АН АрмССР

Поступило 19.I.1966 г.

Ա. Վ. ՄԵՎԻՍ

ԱԴԱՊՏԻՎ ՍԻՍՏԵՄԻ ԿԱՌՈՒՑՄԱՆ ՄԻ ՄԵԹՈԴԻ ՄԱՍԻՆ

Հոդվածում քննարկված են ադապտիվ սիստեմներ կառուցելիս ծագող երկու պրոբլեմներ՝ ինչպիսին պետք է լինի կլասիֆիկացնող ֆունկցիան և ինչպես պետք է ընթանա ուսուցման պրոցեսը։ Բերվում է ադապատիվ սիստեմի մաթեմատիկական մոդելը, որը բաղկացած է ճանաչման և ուսուցման սխեմաներից։ Տրվում է կլասիֆիկացնող ֆունկցիան կազմելու մեթոդը։

ЛИТЕРАТУРА

1. Уидроу Б. Зарубежная радиоэлектроника, 9, 1965.
2. Chow C. K. JRE Trans. El. Comp. vol. 6, 12, 1957.
3. Chow C. K. JEEE Trans. El. Comp. vol. 14, 1, 1965.
4. Widrow B. and Hoff M. E. JRE Wescon Conv. Rec. Pt. 4, pp. 96—101, 1960.
5. Fukunaga K., Ito T. JEEE Trans. El. Comp. vol. 14, 1, 1965.