

О нелинейной сложности нейронных схем Мак-Каллока-Питтса

В. С. Половников

Как правило для параллельных вычислений выполняемых с помощью нейронных сетей важны два аспекта сложности: скорость вычисления (глубина нейросети) и объем необходимой памяти (количество нейронов в сети). Автором рассмотрен класс дискретных нейронных сетей прямого распространения Мак-Каллока-Питтса и изучен соответствующий подкласс нейронных схем без памяти, для которого удалось получить оценки нелинейной сложности (аналога сложностных параметров нейросетей). В работе доказывается, что любую нейронную схему Мак-Каллока-Питтса можно представить схемой нелинейной глубины не более двух, оценена функция Шеннона нелинейной сложности схем. Так же показано, что асимптотика функции Шеннона достигается на схемах нелинейной глубины два.

Рассмотрим представление произвольной нейронной сети [1] модели Мак-Каллока-Питтса [2] в виде схемы из следующих функциональных элементов [4]: «константа» (рис. 1а), «сумматор» (рис. 1б), «умножение на константу» (рис. 1в), « θ -функция» (рис. 1г). Для построения схемы заменим в нейросети веса, приписанные входным сигналам нейронов, на элементы «умножение на константу», а сдвиг h , представленный постоянным входом, на соответствующий элемент «константа».

Ввиду того, что в модели нейрона, схематически представленной на рис. 2, помимо весов входов есть сумматор и функция активации θ , которые присутствуют и в элементном базисе (рис. 1), то получится корректно определенная нейронная схема без памяти (см. работы [5, 6]).

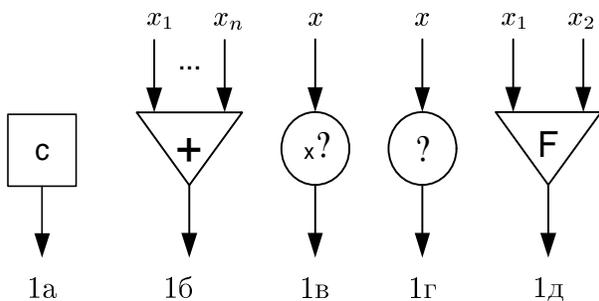


Рис. 1. Элементный базис.

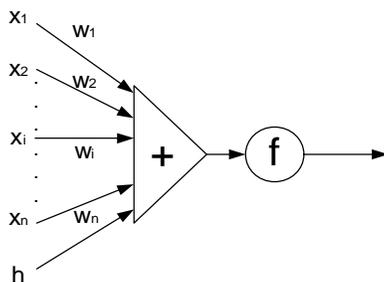


Рис. 2. Математическая модель нейрона.

В этой схеме не будет использован элемент F (рис. 1д), выходным элементом схемы обязательно будет элемент θ -функция. Если мы откажемся от последнего условия, то есть будем строить нейронные схемы без элемента F с произвольным выходным элементом, то реализуемая функция будет представлять из себя линейную комбинацию нейросетевых функций модели Мак-Каллока-Питтса и линейных функций.

Ввиду близости моделей, нейронные схемы без памяти, полученные без использования элемента F по правилам, описанным в [5, 6], будем называть *нейронными схемами Мак-Каллока-Питтса*.

Дадим обозначения понятиям, введенным в [5, 6]:

L — множество линейных функций,

PL — множество кусочно-линейных функций,

PC — множество кусочно-постоянных функций.

Определим множество *кусочно-параллельных функций* следующим выражением:

$$PP = \{f | f = f_c + f_l, f_c \in PC, f_l \in L\}.$$

Множество функций, реализуемых нейронными схемами Мак-Каллока-Питтса обозначим \mathcal{L}' .

Лемма 1. Пусть функция $f(\bar{x}) \in PL$, тогда функция $g(\bar{x}) = \theta(f(\bar{x}))$ — кусочно постоянна.

Доказательство. Действительно, пусть f задана гиперплоскостями l_1, \dots, l_k , делящих \mathbb{R}^n на классы эквивалентности R_1, \dots, R_s , на которых она линейна и для всех $j = 1, \dots, s$ выполнено: при $\bar{x} \in R_j$ $f(\bar{x}) = f|_{R_j} = \bar{b}_j \cdot \bar{x} + d_j$, где $\bar{b}_j \in \mathbb{R}^n, d_j \in \mathbb{R}$. Тогда, очевидно, на классах $R_j, j = 1, \dots, s$, функция $g(\bar{x}) = \theta(\bar{b}_j \cdot \bar{x} + d_j)$, то есть равна 1, при $\bar{b}_j \cdot \bar{x} + d_j \geq 0$ и равна 0 иначе. Таким образом, ввиду линейности условий, функция $g(\bar{x})$ является кусочно-постоянной, принимает значения 0 или 1 и задается следующим набором гиперплоскостей: l_1, \dots, l_k и теми $l'_j = \{\bar{x} | \bar{b}_j \cdot \bar{x} + d_j = 0\}$, где $\bar{b}_j \neq \bar{0}, j = 1, \dots, s$. Лемма доказана.

Теорема 1. Множество функций, реализуемых нейронными схемами Мак-Каллока-Питтса, совпадает с множеством всех кусочно-параллельных функций. И для любой кусочно-параллельной функции существует нейронная схема Мак-Каллока-Питтса нелинейной глубины не более двух.

Доказательство. Докажем включение $\mathcal{L}' \subseteq PP$.

Пусть (S, f) нейронная схема Мак-Каллока-Питтса, из строения схемы легко заключить, что функция f имеет следующий вид:

$$f(\bar{x}) = c_1 f_1(\bar{x}) + \dots + c_p f_p(\bar{x}) + f_l, \quad (1)$$

где p — некоторое натуральное число (может быть нулевым), $c_i \in \mathbb{R}, i = 1, \dots, p, f_l \in L$. Функции f_1, \dots, f_p являются кусочно-постоянными по лемме 1, так как они образованы подсхемами $(S_1, f_1), \dots, (S_p, f_p)$ с выходными элементами θ , а функции f_1, \dots, f_p являются

кусочно-линейными по результатам работ [5, 6]. По определению кусочно-параллельных функций, ввиду линейности пространства PC и выражения (1), следует, что $f \in PP$.

Докажем обратное включение $\mathcal{L}' \supseteq PP$.

Любая линейная функция $u(\bar{x}) = \bar{x} \cdot \bar{a} + c$, $\bar{a} = (a_1, \dots, a_n)$, может быть реализована нейронной схемой без памяти как показано на рис. 3.

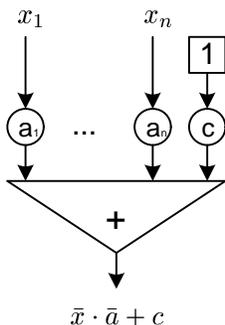


Рис. 3.

Пусть f — кусочно-параллельная функция. По определению, существуют $f_c \in PC$ и $f_l \in L$, что $f = f_c + f_l$. Если мы сможем построить нейронную схему функции f_c , используя лишь сумматор, умножитель на константу и функцию θ , то мы сможем реализовать схему без использования элемента F и для функции f . Действительно, построив схему для линейной функции f_l согласно рис. 3, и подав на входы сумматора выходы схем для функций f_c и f_l , мы получим искомую схему Мак-Каллока-Питтса для функции f . Покажем теперь, как построить схему Мак-Каллока-Питтса для произвольной функции $f_c \in PC$. Пусть кусочно-постоянная функция f_c задана гиперплоскостями $l_i : \bar{a}_i \cdot \bar{x} + c_i = 0$, $i = 1, \dots, k$, которые разбивают \mathbb{R}^n на классы эквивалентности R_1, \dots, R_s , что для некоторых действительных d_1, \dots, d_s при $\bar{x} \in R_j$ выполнено $f(\bar{x}) = f_{R_j} = d_j$, $j = 1, \dots, s$. Тогда будем строить схему для функции f_c по формуле аналогичной представленной в статье [6]:

$$f_c(\bar{x}) = \sum_{j=1}^s d_j \theta \left(\sum_{i=1}^k \chi(\operatorname{sgn}(\bar{a}_i \cdot \bar{x} + c_i) = \sigma_j^i) - k \right), \quad (2)$$

а формула построения схемы для функции f будет иметь вид

$$f(\bar{x}) = \sum_{j=1}^s d_j \theta \left(\sum_{i=1}^k \chi(\operatorname{sgn}(\bar{a}_i \cdot \bar{x} + c_i) = \sigma_j^i) - k \right) + f_i. \quad (3)$$

где σ_j^i — i -я компонента вектора сигнатуры класса R_j , $i = 1, \dots, k$, $j = 1, \dots, s$, а обозначение χ раскрывается по следующим формулам:

$$\chi(\operatorname{sgn}(\bar{a}_i \cdot \bar{x} + c_i) = 1) = \theta'(\bar{a}_i \cdot \bar{x} + c_i) = 1 - \theta(-\bar{a}_i \cdot \bar{x} - c_i), \quad (4)$$

$$\chi(\operatorname{sgn}(\bar{a}_i \cdot \bar{x} + c_i) = 0) = \theta(\bar{a}_i \cdot \bar{x} + c_i) + \theta(-\bar{a}_i \cdot \bar{x} - c_i) - 1, \quad (5)$$

$$\chi(\operatorname{sgn}(\bar{a}_i \cdot \bar{x} + c_i) = -1) = 1 - \theta(\bar{a}_i \cdot \bar{x} + c_i). \quad (6)$$

Как видно из рис. 4–6 (изображения соответствующих нейронных схем для формул 4–6), схемная реализация χ имеет нелинейную глубину один и не содержит элементов F . Теорема доказана.

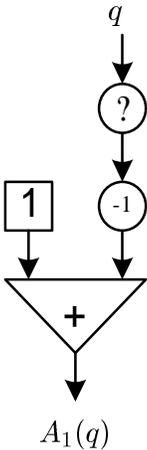


Рис. 4.

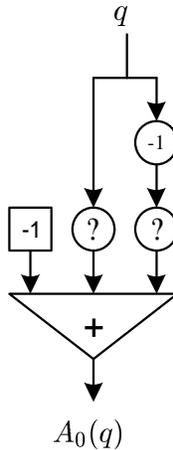


Рис. 5.

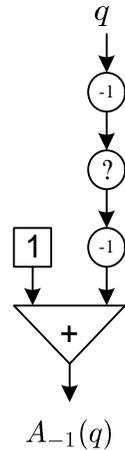


Рис. 6.

Пусть $f(x_1, \dots, x_n)$ — произвольная кусочно-параллельная функция в \mathbb{R}^n , заданная k гиперплоскостями. Согласно теореме 1, существует (S, f) — нейронная схема Мак-Каллока-Питтса нелинейной глубины два, реализующая f . За $Z_{MP}(S)$ обозначим число элементов θ в схеме (S, f) , то есть *нелинейную сложность* схемы (S, f) .

Соответственно определим сложность реализации функции схемой Мак-Каллока-Питтса

$$Z_{MP}(f) = \min_{S \text{ реализует } f} Z_{MP}(S),$$

где минимум берется по всем нейронным схемам Мак-Каллока-Питтса, реализующим f .

$$\text{Функция Шеннона } Z_{MP}(k) = \max_{f \in PP: f \text{ задана } k \text{ гиперплоскостями}} Z_{MP}(f).$$

Пусть $H(n, k)$ — максимальное число классов эквивалентности, на которые k гиперплоскостей могут разбить \mathbb{R}^n .

Лемма 2. Для всех натуральных n и k верно $H(1, k) = 2k + 1$, $H(n, 1) = 3$. А для $n \geq 2$ и $k \geq 2$ выполнено неравенство

$$H(n, k) \leq H(n, k - 1) + 2H(n - 1, k - 1).$$

Доказательство. 1) В \mathbb{R}^1 гиперплоскостями являются точки. Они разбивают прямую на максимальное число классов эквивалентности, когда не совпадают. Классами являются: 2 луча, k точек и $k - 1$ отрезок. Таким образом, $H(1, k) = 2k + 1$.

2) В \mathbb{R}^n одна гиперплоскость разбивает все пространство на 3 класса эквивалентности: 2 полупространства и саму эту гиперплоскость. Получаем $H(n, 1) = 3$.

3) Гиперплоскости l_1, \dots, l_{k-1} разбивают \mathbb{R}^n на классы эквивалентности R_1, \dots, R_s , $s \leq H(n, k - 1)$. Рассмотрим пересечения гиперплоскостей l_i с l_k , $i = 1, \dots, k - 1$. Известно, если две гиперплоскости в \mathbb{R}^n пересекаются и не совпадают, то пересечением является линейное пространство размерности $n - 2$. Значит, пересечения образуют $k' \leq k - 1$ гиперплоскостей в \mathbb{R}^{n-1} , которые разбивают l_k на $s' \leq H(n - 1, k')$ классов эквивалентности. Ввиду очевидной монотонности $H(n, k)$ по k , этих классов не более чем $H(n - 1, k - 1)$ и каждый из них является пересечением одного из R_1, \dots, R_s с l_k . Учитывая, что каждый из s' классов, который гиперплоскость l_k пересекает, делится не более чем на 3 класса, получаем доказываемое неравенство. Лемма доказана.

Будем говорить, что k гиперплоскостей находятся в *общем положении* в \mathbb{R}^n в следующих случаях:

- 1) при $n = 1$, если все точки (гиперплоскости в \mathbb{R}^1 различны);
- 2) при $k \leq n$, $n \geq 2$, если все k векторов нормалей к гиперплоскостям линейно независимы;
- 3) при $k > n$, $n \geq 2$, если
 - а) любые n векторов нормалей к гиперплоскостям линейно независимы,
 - б) для каждой гиперплоскости l остальные $k - 1$ гиперплоскости разбивают \mathbb{R}^n на классы эквивалентности, целиком не лежащие на l .

Легко убедиться, что произвольные k гиперплоскостей, находящиеся в общем положении в пространстве \mathbb{R}^n , разбивают пространство на одинаковое число классов эквивалентности. Обозначим это число $H'(n, k)$, по определению $H'(n, k) \leq H(n, k)$.

Лемма 3. Для всех натуральных n и k верно $H'(1, k) = 2k + 1$, при $n \geq k$ выполнено $H'(n, k) = 3^k$. А для $n \geq 2$ и $k > n$ имеет место следующее соотношение:

$$H'(n, k) = H'(n, k - 1) + 2H'(n - 1, k - 1). \quad (7)$$

Доказательство. Пусть гиперплоскости l_1, \dots, l_k находятся в общем положении в \mathbb{R}^n .

1) Так как точки, находящиеся в общем положении в \mathbb{R}^1 , различны, то $H'(1, k) = 2k + 1$.

2) Если $n \geq k$ и все k векторов нормалей к гиперплоскостям линейно независимы, то каждая последующая гиперплоскость делит все предыдущие классы эквивалентности на три. Таким образом, из $H'(1, 1) = 3$ следует $H'(n, k) = 3^k$.

3) Рассмотрим гиперплоскость l_k . В силу того, что любые два вектора нормали линейно независимы, l_1, \dots, l_{k-1} пересекают l_k по линейным пространствам размерности $n - 2$. Обозначим их l'_1, \dots, l'_{k-1} — гиперплоскости в $l_k \cong \mathbb{R}^{n-1}$. Пусть $\bar{a}^1, \dots, \bar{a}^k$ — суть векторы нормали к гиперплоскостям l_1, \dots, l_k соответственно. Введем систему координат $\bar{e}^1, \dots, \bar{e}^n$, в которой вектор \bar{a}_k имеет вид $(0, \dots, 0, a_n^k)$, а остальные \bar{a}^i произвольны, $\bar{a}^i = (a_1^i, \dots, a_n^i)$, $i = 1, \dots, k - 1$. Тогда гиперплоскости l'_1, \dots, l'_{k-1} в \mathbb{R}^{n-1} имеют векторы нормали $\bar{a}^i = (a_1^i, \dots, a_{n-1}^i)$,

$i = 1, \dots, k - 1$. Причем любые $n - 1$ из этих векторов линейно независимы (действительно, если для некоторых i_1, \dots, i_{n-1} векторы $\bar{a}_1^{i_1}, \dots, \bar{a}_1^{i_{n-1}}$ были бы линейно зависимы, то и соответствующие им векторы $\bar{a}_1^{i_1}, \dots, \bar{a}_1^{i_{n-1}}$ вместе с вектором \bar{a}_k составили бы линейно зависимую систему, что противоречит общему положению l_1, \dots, l_k в \mathbb{R}^n). Предположим, что какая-то из l'_i содержит целиком класс эквивалентности на которые $l'_1, \dots, l'_{i-1}, l'_{i+1}, \dots, l'_{k-1}$ разбивают \mathbb{R}^{n-1} , тогда и l_i содержит этот класс (так как $l'_i = l_i \cap l_k$), что также противоречит общему положению исходных гиперплоскостей. Таким образом, гиперплоскости l'_1, \dots, l'_{k-1} находятся в общем положении.

Очевидно, что и l_1, \dots, l_{k-1} находятся в общем положении, а гиперплоскость l_k разбивает ровно на три части каждый из пересекаемых ей $H'(n - 1, k - 1)$ из общего числа $H'(n, k - 1)$ классов эквивалентности, на которые l_1, \dots, l_{k-1} разбивают \mathbb{R}^n (иначе существовал бы класс целиком лежащий на l_k). Это доказывает верность формулы. Лемма доказана.

Замечание. Отметим, что соотношение (7) выполнено при любых натуральных $n \geq 2$ и $k \geq 2$. Действительно, по доказанной лемме 3, $H'(n, k) = 3^k$, что удовлетворяет уравнению (7) при указанных значениях n и k .

Следствие 1. k гиперплоскостей, находящиеся в общем положении в \mathbb{R}^n , разбивают пространство на максимально возможное количество классов эквивалентности. То есть для всех натуральных n и k верно

$$H'(n, k) = H(n, k).$$

Доказательство. Граничные условия совпадают $H'(1, k) = H(1, k) = 2k + 1$ и $H'(n, 1) = H(n, 1) = 3$, и выполнено соотношение $H'(n, k) = H'(n, k - 1) + 2H'(n - 1, k - 1)$ для всех натуральных $n \geq 2$ и $k \geq 2$.

Пусть $H(n, k) = H'(n, k) + h(n, k)$, $h(n, k) \geq 0$ и, согласно лемме 2, $H(n, k) \leq H(n, k - 1) + 2H(n - 1, k - 1)$. Тогда $h(1, k) = 0$, $h(n, 1) = 0$, $0 \leq h(n, k) \leq h(n, k - 1) + 2h(n - 1, k - 1)$. Откуда, очевидно, получаем $h(n, k) \equiv 0$ и $H(n, k) = H'(n, k)$. Следствие доказано.

Лемма 4. При $k > n$ выполнено

$$H(n, k) = 3 \sum_{i=0}^{n-1} C_{k-1}^i 2^i + 2^n \sum_{j=n-1}^{k-2} C_j^{n-1}.$$

Доказательство.

Введем обозначение $\bar{H}(k) = \begin{pmatrix} H(n, k) \\ H(n-1, k) \\ \vdots \\ H(1, k) \end{pmatrix},$

тогда $\bar{H}(1) = (3, 3, \dots, 3)^T$. Учитывая (7) и следствие леммы 3 получаем

$$\bar{H}(k) = A\bar{H}(k-1) + \bar{c},$$

где $A = \underbrace{\begin{pmatrix} 1 & 2 & 0 & \dots & 0 & 0 \\ 0 & 1 & 2 & \dots & 0 & 0 \\ & & \ddots & \ddots & & \\ 0 & 0 & \dots & 1 & 2 & 0 \\ 0 & 0 & \dots & 0 & 1 & 2 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}}_n, \quad \bar{c} = \left. \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 2 \end{pmatrix} \right\} n$

Откуда получаем

$$\begin{aligned} \bar{H}(k) &= A(A\bar{H}(k-2) + \bar{c}) + \bar{c} = A^2\bar{H}(k-2) + A\bar{c} + \bar{c} = \\ &\dots = A^{k-1}\bar{H}(1) + A^{k-2}\bar{c} + \dots + A\bar{c} + \bar{c}. \end{aligned} \quad (8)$$

Нас интересует $H(n, k)$, то есть первый элемент вектора $\bar{H}(k)$. Значит, в соответствии с выражением (8), видом векторов $\bar{H}(1)$ и \bar{c} , нужно найти первую строку матрицы A^{k-1} и n -й элемент матриц A^{k-2}, \dots, A .

Пусть

$$B = \begin{pmatrix} 0 & 2 & 0 & \dots & 0 & 0 \\ 0 & 0 & 2 & \dots & 0 & 0 \\ & & \ddots & \ddots & & \\ 0 & 0 & \dots & 0 & 2 & 0 \\ 0 & 0 & \dots & 0 & 0 & 2 \\ 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}, \quad E = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ & & \ddots & \ddots & & \\ 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix},$$

$$\text{тогда } A^j = (E + B)^j = \sum_{i=0}^j C_j^i E^{j-i} B^i = \sum_{i=0}^j C_j^i B^i.$$

$$\text{Так как } B^i = \begin{pmatrix} \overbrace{0 \dots 0}^i & 2^i & 0 & \dots & 0 & 0 \\ 0 \dots 0 & 0 & 2^i & & 0 & 0 \\ \vdots & & \ddots & \ddots & \vdots & \\ 0 & \dots & 0 & 2^i & 0 & \\ 0 & \dots & 0 & 0 & 2^i & \\ 0 & \dots & 0 & 0 & 0 & \\ \vdots & & \vdots & \vdots & \vdots & \\ 0 & \dots & 0 & 0 & 0 & \end{pmatrix}, \text{ то при } i \leq n \text{ матрица}$$

$B^i = 0$ и первый элемент $A^{k-1} \bar{H}(1)$ есть

$$H_{k-1}(n, k) = 3 \sum_{i=0}^{n-1} C_{k-1}^i 2^i, \quad (9)$$

а первый элемент $A^i \bar{c}$, при $i = n-1, \dots, k-2$ есть

$$H_i(n, k) = C_i^{n-1} 2^n. \quad (10)$$

Очевидно, при $i = 1, \dots, n-1$, $A^i \bar{c}$ обращается в ноль.

Подставляя (9) и (10) в (8) получаем первый элемент столбца $\bar{H}(k)$

$$H(n, k) = \sum_{i=n-1}^{k-1} H_i(n, k) = 3 \sum_{i=0}^{n-1} C_{k-1}^i 2^i + 2^n \sum_{j=n-1}^{k-2} C_j^{n-1}.$$

Лемма доказана.

Следствие 2. Асимптотика $H(n, k) \sim \frac{2^n}{n!} k^n$, при $n = \text{const}$, $k \rightarrow \infty$.

Доказательство. Введем обозначения

$$H_1(n, k) = 3 \sum_{i=0}^{n-1} C_{k-1}^i 2^i, \quad H_2(n, k) = 2^n \sum_{j=n-1}^{k-2} C_j^{n-1}.$$

Тогда при $n = \text{const}$, $k \rightarrow \infty$

$$H_1(n, k) \sim 3 \cdot 2^{n-1} \cdot C_{k-1}^{n-1} \sim \frac{3 \cdot 2^{n-1}}{(n-1)!} k^{n-1}.$$

Преобразуем $H_2(n, k)$ к виду

$$H_2(n, k) = \frac{2^n}{(n-1)!} \sum_{i=0}^{k-n-1} \frac{(i+n-1)!}{i!} = \frac{2^n}{(n-1)!} a_{k-n-1},$$

$$\text{где } a_p = \sum_{i=0}^p \frac{(i+n-1)!}{i!}.$$

Легко видеть, что a_p удовлетворяет следующему рекуррентному соотношению:

$$\begin{cases} a_p + 1 = a_p + Q_{n-1}(p), \\ a_{-1} = 0, \end{cases}$$

$$\text{где } Q_{n-1} = \frac{(p+n)!}{(p+1)!} - \text{многочлен степени } n-1.$$

Известно (см. [7]), что решением такого уравнения будет полином степени n , то есть $a_p = R_n(p) = \gamma p^n + \dots$

Вычислим коэффициент γ при старшем члене $R_n(p)$. Очевидно, так как старший коэффициент $Q_{n-1}(p)$ равен единице,

$$\gamma n = \lim_{p \rightarrow \infty} \frac{R_n(p+1) - R_n(p)}{p^{n-1}} = \lim_{p \rightarrow \infty} \frac{Q_{n-1}(p)}{p^{n-1}} = 1, \quad \gamma = \frac{1}{n}.$$

Значит,

$$a_p \sim \frac{p^n}{n}, \quad \text{а } H_2(n, k) \sim \frac{2^n}{(n-1)!} \frac{(k-n-1)^n}{n} \sim \frac{2^n}{n!} k^n,$$

при $n = \text{const}$, $k \rightarrow \infty$.

В итоге при $n = \text{const}$, $k \rightarrow \infty$,

$$H(n, k) = H_1(n, k) + H_2(n, k) \sim \frac{3 \cdot 2^{n-1}}{(n-1)!} k^{n-1} + \frac{2^n}{n!} k^n \sim \frac{2^n}{n!} k^n.$$

Следствие доказано.

Теорема 2. Для функции Шеннона $Z_{MP}(k)$ верна оценка

1) при $k \leq n$:

$$3^k - 1 \leq Z_{MP}(k) \leq 3^k + 2k;$$

2) при $k > n$:

$$\begin{aligned} 3 \sum_{i=0}^{n-1} C_{k-1}^i 2^i + 2^n \sum_{j=n-1}^{k-2} C_j^{n-1} - 1 &\leq Z_{MP}(k) \leq \\ &\leq 3 \sum_{i=0}^{n-1} C_{k-1}^i 2^i + 2^n \sum_{j=n-1}^{k-2} C_j^{n-1} + 2k. \end{aligned}$$

Доказательство.

1. Реализуем функцию, для которой количество классов эквивалентности $s = H(n, k)$ по формуле (3). Нам потребуются $2k$ элементов $\theta(\bar{a}_i \cdot \bar{x} + c_i)$, $\theta(-\bar{a}_i \cdot \bar{x} - c_i)$ на первом нелинейном ярусе для реализации всевозможных $\chi(\text{sgn}(\bar{a}_i \cdot \bar{x} + c_i) = \sigma_j^i)$. Далее, на втором нелинейном ярусе необходимы $s = H(n, k)$ элементов θ .

Таким образом мы получили оценку сверху:

$$Z_{MP}(k) \leq s + 2k = H(n, k) + 2k.$$

Учитывая результат леммы 4, запишем оценку в виде:

$$Z_{MP}(k) \leq \begin{cases} 3 \sum_{i=0}^{n-1} C_{k-1}^i 2^i + \\ \quad 2^n \sum_{j=n-1}^{k-2} C_j^{n-1} + 2k, & \text{при } k > n, \\ 3^k + 2k, & \text{при } k \leq n. \end{cases} \quad (11)$$

2. Рассмотрим кусочно-постоянную функцию f , для которой задающие ее гиперплоскости находятся в общем положении. Пусть она

принимает значения e^j на j -м классе эквивалентности $j = 1, \dots, s$, $s = H(n, k)$, где e — некое трансцендентное число [8]. Рассмотрим так называемый внешний нелинейный слой минимальной по нелинейной сложности схемы (S, f) , то есть множество элементов θ , являющихся последними в путях от входов к выходу. Занумеруем их каким-нибудь способом и через $\tau(\bar{x})$ обозначим вектор значений элементов внешнего нелинейного слоя (вектор из нулей и единиц) когда на вход схемы подается \bar{x} . Очевидно, что при фиксированном наборе τ схема S реализует некоторую определенную линейную функцию, независимо от значений остальных нелинейных элементов (не внешнего нелинейного слоя). Пусть $\tau(\bar{x})$ состоит из t элементов, тогда схема S может реализовать соответственно не более 2^t различных линейных функций. Каждый из элементов внешнего нелинейного слоя вносит, ввиду отсутствия нелинейных элементов ниже этого слоя, некий аддитивный вклад $c_i \in \mathbb{R}$ в реализуемую схемой S функцию, причем только тогда, когда соответствующая компонента вектора τ принимает значение 1, $i = 1, \dots, t$ и вносит нулевой вклад иначе. Итак, S реализует функцию $f(\bar{x}) = \tau(\bar{x}) \cdot \bar{c} + f_l(\bar{x})$, где $\bar{c} = (c_1, \dots, c_t)$. Линейная составляющая f_l постоянна $f_l = c_0 \in \mathbb{R}$ ввиду того, что реализуемая функция кусочно-постоянна. Действительно, пусть $f_l(\bar{x}) = \bar{a} \cdot \bar{x} + c_0$. Рассмотрим такие векторы $\bar{x}^{(0)}, \bar{x}^{(1)}, \dots, \bar{x}^{(n)} \in \mathbb{R}^n$ из какого-либо n -мерного класса R_j (не лежащем целиком ни на одной гиперплоскости), что каждый вектор $\bar{x}^{(i)}$ отличается от вектора $\bar{x}^{(0)}$ лишь в i -й компоненте, $i = 1, \dots, n$ и $\tau(\bar{x}^{(1)}) = \dots = \tau(\bar{x}^{(n)})$. Такие точки обязательно существуют ввиду конечности всевозможных различных векторов τ (их не более 2^t) и бесконечности класса R_j . Рассмотрим уравнения соответствующие схеме (S, f) когда на ее вход подаются $\bar{x}^{(i)}$:

$$f(\bar{x}^{(i)}) = \tau(\bar{x}^{(i)}) \cdot \bar{c} + \bar{a} \cdot \bar{x}^{(i)} + c_0 = e^j, \quad (12)$$

так как $\bar{x}^{(i)} \in R_j$ и $f|_{R_j} = e^j$, $i = 0, \dots, n$. Исходя из того, что вектор τ постоянен в точках $\bar{x}^{(0)}, \dots, \bar{x}^{(n)}$, а значит скалярное произведение $\tau(\bar{x}^{(i)}) \cdot \bar{c} = d = \text{const}$ и уравнения (12) являются линейной системой из $n + 1$ уравнений с $n + 1$ неизвестной: $\bar{a} = (a_1, \dots, a_n)$ и c_0 . Вычитая из уравнения для $\bar{x}^{(i)}$ уравнение для $\bar{x}^{(0)}$ получаем, что $a_i = 0$, $i = 1, \dots, n$. Таким образом, $a_1 = \dots = a_n = 0$ и $f_l(\bar{x}) = c_0$.

Для каждого значения функции e^1, \dots, e^s выберем по одному вектору τ , на котором это значение достигается (заметим, что все выбранные s векторов τ_1, \dots, τ_s различны). Учитывая, что $f_l(\tilde{x}) = c_0$, можем действие схемы (S, f) , когда на внешнем нелинейном слое формируются векторы τ_1, \dots, τ_s , записать в матричном виде следующим образом:

$$\begin{pmatrix} 1 & a_{11} & \dots & a_{1t} \\ 1 & a_{21} & \dots & a_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{s1} & \dots & a_{st} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_t \end{pmatrix} = \begin{pmatrix} e^1 \\ e^2 \\ \vdots \\ e^s \end{pmatrix}, \quad (13)$$

где векторы $\tau_j = (a_{j1}, \dots, a_{jt}) \in \{0, 1\}^t$, $j = 1, \dots, s$. Будем рассматривать бинарные векторы $(1, a_{j1}, \dots, a_{jt})$, $j = 1, \dots, s$, как элементы векторного пространства \mathbb{Q}^{t+1} .

Предположим, что $s \geq t + 2$, тогда любые $t + 2$ вектора, в том числе и первые $(1, a_{j1}, \dots, a_{jt})$, $j = 1, \dots, t + 2$ линейно зависимы в \mathbb{Q}^{t+1} , то есть существуют коэффициенты $\alpha_1, \dots, \alpha_{t+2} \in \mathbb{Q}$, не все обращающиеся в 0, что

$$\alpha_1(1, a_{11}, \dots, a_{1t}) + \dots + \alpha_{t+2}(1, a_{t+2;1}, \dots, a_{t+2;t}) = (0, \dots, 0). \quad (14)$$

Домножим скалярно обе части выражения (14) на вектор-столбец $(c_0, c_1, \dots, c_t)^T$, учитывая (13), получим равенство

$$\alpha_1 e^1 + \dots + \alpha_{t+2} e^{t+2} = 0. \quad (15)$$

Равенство (15) означает, что трансцендентное число e является корнем уравнения с рациональными коэффициентами, то есть является алгебраическим, что противоречит определению трансцендентного числа.

Следовательно, $s \leq t + 1$. Учитывая, что t — количество нелинейных элементов на внешнем слое в минимальной схеме (S, f) , выполнена оценка

$$Z_{MP}(k) \geq Z_{MP}(f) \geq t \geq s - 1 = H(n, k) - 1.$$

По лемме 4 ее можно переписать в следующем виде:

$$Z_{MP}(k) \geq \begin{cases} 3 \sum_{i=0}^{n-1} C_{k-1}^i 2^i + \\ 2^n \sum_{j=n-1}^{k-2} C_j^{n-1} - 1, & \text{при } k > n, \\ 3^k - 1, & \text{при } k \leq n. \end{cases} \quad (16)$$

Объединяя оценки (11) и (16) получаем утверждение теоремы. Теорема доказана.

Учитывая следствие леммы 4, легко получить асимптотику.

Следствие 3. *Для функции Шеннона $Z_{MP}(k)$ верна асимптотика*

$$Z_{MP}(k) \sim \frac{2^n}{n!} k^n, \quad k \rightarrow \infty, \quad n = \text{const}, \quad n > 1.$$

Автор выражает благодарность научному руководителю Часовских А. А.

Список литературы

- [1] Хайкин С. Нейронные сети: полный курс / 2-е издание. Вильямс, 2006.
- [2] McCulloch W. S., Pitts W. A logical calculus of the ideas immanent in nervous activity // Bulletin of Mathematical Biophysics. 1943. Vol. 5. P. 115–133.
- [3] Кудрявцев В. Б. Функциональные системы. М.: Изд-во МГУ, 1982.
- [4] Лупанов О. Б. Асимптотические оценки сложности управляющих систем. М.: МГУ, 1984.
- [5] Половников В. С. О некоторых характеристиках нейронных схем // Вестн. Моск. Ун-та. Сер. 1: Математика. Механика. 2004. № 5. 65–67.
- [6] Половников В. С. О некоторых характеристиках нейронных схем // Интеллектуальные системы. Т. 8. Вып. 1–4. 2004. 121–145.
- [7] Кофман А. Введение в прикладную комбинаторику. М.: Наука, 1975.
- [8] Гельфонд А. О. Трансцендентные и алгебраические числа / Изд. 2. М.: КомКнига, 2006.

