Особенности нейронных схем Мак-Каллока – Питтса над полем рациональных чисел

В. С. Половников

В диссертации автора «Об оптимизации структурной реализации нейронных сетей» нейронные схемы рассматривались над полем действительных чисел. Большинство результатов, относящихся к схемам Мак-Каллока – Питтса, дословно переносится на нейронные схемы над полем рациональных чисел, однако для доказательства одной из теорем существенно использовалось понятие трансцендентного числа. Тем не менее результат теоремы для нейронных схем над полем рациональных чисел сохраняется. Новое доказательство приводится в данной работе.

Ключевые слова: нейронные схемы, функция Шеннона.

Кусочно-линейные (PL^n) и кусочно-параллельные (PP^n) функции в $\mathbb{Q}^n$ определяются аналогично функциям в $\mathbb{R}^n$ (см. [1, 2]).

Если гиперплоскости $l_1, \ldots, l_k$ находятся в общем положении ([3]), то они разбивают $\mathbb{Q}^n$ на максимально возможное количество $H(n, k)$ классов эквивалентности.

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

Пусть $f(x_1, \ldots, x_n)$ — произвольная кусочно-параллельнейная функция в $\mathbb{Q}^n$, заданная k гиперплоскостями. Тогда, 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.

Функция Шеннона $Z_{MP}(n,k) = \max\limits_{f\in PP^n: f \text{задана } k \text{ гиперплоскостями}} Z_{MP}(f).$

Теорема 1. *Функция Шеннона $Z_{MP}(n,k)$ удовлетворяет неравенству:*

$$H(n,k) - 1 \leqslant Z_{MP}(n,k) \leqslant H(n,k) - 2k.$$

Для доказательства теоремы нам потребуется следующая лемма:

Лемма 1. *Для любого натурального $s \geqslant 2$ и рационального вектора $\bar{b} = (b_1, \ldots, b_s) \in Q^s$ не существует рационального вектора $\bar{c} = (c_1, \ldots, c_t) \in Q^t$, $t < s$ и бинарной матрицы A, что*

$$A \cdot \bar{c} = \bar{b}. \tag{1}$$

Доказательство. Доказательство проведем по индукции по s. При $s = 2$ очевидно, любые 2 различные ненулевые рациональные числа обладают необходимым свойством. Пусть доказано для $s-1$ и $\bar{b}$ — именно тот вектор длины $s-1$, для которого не существует более короткого вектора $\bar{c}$ и бинарной матрицы A, что выполнено (1). Заметим, что вектор длины $s-1$ всегда существует (возьмем в качестве $\bar{c}$, например, сам вектор $\bar{b}$, а матрицу A единичную). Рассмотрим теперь всевозможные бинарные матрицы A размера $(s-1)$ x $(s-1)$ и множество решений (1) как уравнения относительно вектора $\bar{c} \in Q^{s-1}$ при фиксированной матрице A и векторе $\bar{b}$. Заметим, что для существования решения необходимо, чтобы матрица A была невырождена (и решение при этом будет единственно). Действительно, если бы матрица A была вырожденной, то существовал бы столбец, линейно выражаемый с рациональными коэффициентами через другие столбцы. Обозначим элемент j-й строки, i-го столбца матрицы A через a_i^j, элементы векторов $\bar{b}$ и $\bar{c}$ через b_j и c_i соответственно, $i = 1, \ldots, s-1$,

$j = 1, \dots, s-1$. Тогда, в виду вырожденности A, не теряя общности, считаем что существуют $\alpha_i \in \mathbb{Q}, i = 1, \dots, s-2$, что выполнено

$$a_{s-1}^j = \sum_{i=1}^{s-2} \alpha_i a_i^j. \qquad (2)$$

Подставим в (1) выражение для a_{s-1}^j полученное в (2). Тогда для всех $j = 1, \dots, s-1$ верно

$$b_j = \sum_{i=1}^{s-2} a_i^j c_i + a_{s-1}^j c_{s-1} = \sum_{i=1}^{s-2} a_i^j (c_i + \alpha_i c_{s-1}).$$

Таким образом мы нашли бинарную матрицу A' — матрица A без последнего столбца, а также вектор $\bar{c}' = (c_1 - \alpha_1 c_{s-1}, \dots, c_{s-2} - \alpha_{s-2} c_{s-1})$, такие что $A'\bar{c}' = \bar{b}$. Тем самым мы получили противоречие выбору вектора $\bar{b}$. Итак, матрица A невырождена и для фиксированной матрицы A существует не более одного $\bar{c}$, удовлетворяющего (1) при данном $\bar{b}$. Для каждого из не более чем $2^{(s-1)(s-1)}$ решений найдем всевозможные $b_s = \bar{a} \cdot \bar{c}$, где $\bar{a} \in \{0,1\}^{s-1}$. Получим конечное число «запрещенных» вариантов числа b_s. Следовательно, для вектора длины s можно выбрать произвольное рациональное «незапрещенное» b_s.

Теперь докажем теорему.

Доказательство теоремы. 1. Реализуем произвольную кусочно-параллельную функцию $f : \mathbb{Q}^n \to \mathbb{Q}$, заданную k гиперплоскостями по следующей формуле:

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

где $f = f_c + f_l$, f_c — кусочно-постоянная функция, f_l — линейная функция.

В свою очередь, f_c задана k гиперплоскостями $l_i = \{\bar{x} \in \mathbb{Q}^n : \bar{a}_i \cdot \bar{x} + c_i = 0\}$, $\bar{a}_i \in \mathbb{Q}^n$, $c_i \in \mathbb{Q}, i = 1, \dots, k$.

Гиперплоскости $l_1, \dots, l_k$ разбивают $\mathbb{Q}^n$ на s классов эквивалентности: $R_1, \dots, R_s$, $f|_{R_j} = d_j$, $d_j \in \mathbb{Q}$, $j = 1, \dots, s$, а σ_j^i — i-ый элемент вектора сигнатуры $\bar{\sigma}(R_j)$ j-го класса эквивалентности.

$$\chi(a,b) = \begin{cases} 1, & \text{при } a = b \\ 0, & \text{иначе.} \end{cases}$$

Нам потребуется $2k$ 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 $\chi\left(sgn(\bar{a}_i \cdot \bar{x} + c_i), \sigma_j^i\right)$. Далее, на втором нелинейном ярусе необходимы $s \leqslant H(n,k)$ элементов θ.

Таким образом мы получили оценку сверху:

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

2. Рассмотрим кусочно-постоянную функцию f, для которой задающие ее гиперплоскости находятся в общем положении. Пусть она принимает значения компонент вектора $\bar{b}$ из леммы 1: $b_j \in \mathbb{Q}$ на j-м классе эквивалентности $j = 1, \ldots, s$, $s = H(n,k)$. Рассмотрим так называемый внешний нелинейный слой минимальной по нелинейной сложности схемы (S, f), то есть множество элементов θ последних в путях от входов к выходу. Занумеруем их каким-либо 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 $c_i \in \mathbb{Q}$ в реализуемую схемой S функцию, причем только тогда, когда соответствующая компонента вектора τ принимает значение 1, $i = 1, \ldots, t$ и вносит нулевой вклад иначе. Итак, S реализует функцию $f(\bar{x}) = \tau(\bar{x}) \cdot \bar{c} + f_l(\bar{x})$, где $\bar{c} = (c_1, \ldots, c_t)$. Линейная составляющая $f_l = c_0 \in \mathbb{Q}$ постоянна ввиду того, что реализуемая функция кусочно-постоянна. Действительно, пусть $f_l(\bar{x}) = \bar{a} \cdot \bar{x} + c_0$. Рассмотрим такие векторы $\bar{x}^{(0)}, \bar{x}^{(1)}, \ldots, \bar{x}^{(n)} \in \mathbb{Q}^n$ из какого-либо n-мерного класса R_j (не лежащего целиком ни на одной гиперплоскости), что каждый вектор $\bar{x}^{(i)}$ отличается от вектора $\bar{x}^{(0)}$ лишь в i-й компоненте, $i = 1, \ldots, n$ и $\tau(\bar{x}^{(1)}) = \ldots = \tau(\bar{x}^{(n)})$. 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 $\bar{x}^{(i)}$:

$$f(\bar{x}^{(i)}) = \tau(\bar{x}^{(i)}) \cdot \bar{c} + \bar{a} \cdot \bar{x}^{(i)} + c_0 = b_j, \tag{3}$$

так как $\bar{x}^{(i)} \in R_j$ и $f|_{R_j} = b_j$, $i = 0, \ldots, n$. Так как вектор τ постоянен в точках $\bar{x}^{(0)}, \ldots, \bar{x}^{(n)}$, скалярное произведение $\tau(\bar{x}^{(i)}) \cdot \bar{c} = d = \mathrm{const}$ и уравнения (3) являются линейной системой из $n + 1$ уравнений с $n + 1$ неизвестной: $\bar{a} = (a_1, \ldots, a_n)$ и c_0. Вычитая из уравнения для $\bar{x}^{(i)}$ уравнение для $\bar{x}^{(0)}$ получаем $a_i = 0$, $i = 1, \ldots, n$. Таким образом, $a_1 = \ldots = a_n = 0$ и $f_l(\bar{x}) = c_0$.

Для каждого значения функции $b_1, \ldots, b_s$ выберем по одному вектору τ, на котором это значение достигается (заметим, что все выбранные s векторов $\tau_1, \ldots, \tau_s$ различны). Учитывая, что $f_l(\bar{x}) = c_0$ и на внешнем нелинейном слое формируется векторы $\tau_1, \ldots, \tau_s$, можем действие схемы (S, f) записать в матричном виде следующим образом:

$$\begin{pmatrix} 1 & a_{11} & \ldots & a_{1t} \\ 1 & a_{21} & \ldots & a_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{s1} & \ldots & a_{st} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_t \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{pmatrix}, \tag{4}$$

где векторы $\tau_j = (a_{j1}, \ldots, a_{jt}) \in \{0, 1\}^t$, $j = 1, \ldots, s$. Согласно лемме 1 и выбору чисел $b_1, \ldots, b_s$ верно $t \geqslant s - 1 = H(n, k) - 1$. То есть уже на внешнем слое произвольной реализации функции f нейронной схемой Мак-Каллока – Питтса количество нейронов не менее $H(n, k) - 1$. Теорема доказана.

Следствие 1. *Для функции Шеннона $Z_{MP}(n, k)$ верна асимптотика*

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

Список литературы

[1] Половников В. С. О некоторых характеристиках нейронных схем // Вестн. Моск. Ун-та. Сер. 1: Математика. Механика. — М., 2004. № 5. — С. 65–67.

[2] Половников В. С. О некоторых характеристиках нейронных схем // Интеллектуальные системы. — 2004. Т. 8, вып. 1–4. — С. 121–145.

[3] Половников В. С. О нелинейной сложности нейронных схем Мак-Каллока-Питтса // Интеллектуальные системы. — 2007. Т. 11, вып. 1–4. — С. 261–275.