О сложности взаимной обучаемости почти всех нейронов

А. П. Соколов

Работа посвящена вопросам сложности обучения нейронов. В качестве математической модели нейронов рассматриваются пороговые функции алгебры логики. Рассматривается вопрос сложности взаимной обучаемости пар пороговых функций в большинстве случаев. В качестве средства задания пороговых функций выступают целочисленные линейные формы. В качестве меры сложности процесса обучения принято единичное изменение весового коэффициента линейной формы. Показано, что для почти всех пар пороговых функций от n переменных сложность перестройки одной пороговой функции, заданной линейной формой, в другую с ростом n растет экспоненциально.

Ключевые слова: пороговые функции, сложность обучения нейросетей.

Введение

Пороговые функции алгебры логики являются математической моделью нейронов. Они представляют интерес благодаря своим универсальным вычислительным возможностям, а также благодаря возможности их обучения. Последнее свойство с успехом применяется на практике при решении плохо формализуемых задач.

В качестве средства задания пороговых функций в работе рассматриваются линейные формы вида $x_1w_1 + \ldots + x_nw_n - \sigma$ с целочисленными коэффициентами и свободным членом.

Исследуется сложность преобразования одной пороговой функции, заданной линейной формой, к другой, путем пошагового изменения коэффициентов линейной формы. В качестве меры сложности данного процесса принимается изменение коэффициента или свободного члена линейной формы на единицу. Данный процесс может интерпретироваться как процесс обучения нейрона с пороговой функцией активации.

В работе [3], для характеризации сложности обучения в худшем случае исследовалась шенноновская функция $\rho(n)$. Она говорит о том, сколько минимально достаточно выполнить единичных модификаций исходной линейной формы от n переменных для задания желаемой пороговой функции. Было показано, что при стремлении n к бесконечности величина $\log_2 \rho(n)$ растет по порядку как $n \log_2 n$. В работе [4] был конструктивно построен класс пороговых функций расстояние между которыми было ограниченным заранее заданной величиной. При этом мощность данного класса росла экспоненциально с ростом числа переменных. Также можно построить пример большого класса пороговых функций, расстояние между которыми, наоборот, является большим.

Естественным образом возникает вопрос о том, как ведет себя расстояние между пороговыми функциями в большинстве случаев.

В данной работе показано, что для почти всех пар пороговых функций от n переменных сложность перестройки одной пороговой функции, заданной линейной формой, в другую с ростом n растет экспоненциально.

1. Основные понятия, постановки и результаты

Пусть $U = \{u_1, u_2, \ldots\}$ — счетный алфавит переменных. Каждое из переменных u_i может принимать значения из множества $E_2 = \{0, 1\}$. В дальнейшем во избежание употребления сложных индексов мы будем использовать для обозначения букв алфавита U метасимволы x_i с индексами или без них.

Введем определения линейной формы и пороговой функции. *Линейной формой* назовем функцию вида

$$l_{\vec{w},\sigma}(x_1,\ldots,x_n)=\sum_{i=1}^{n}x_iw_i-\sigma,$$

где w_i и σ суть целые числа при $i=1,\ldots,n$. Вектор $\vec{w}=(w_1,\ldots,w_n)$ называют вектором весовых коэффициентов, а σ — порогом.

Функция $f(x_1,\ldots,x_n):E_2^n\to E_2$, называется *пороговой*, если существует линейная форма $l_{\vec{w},\sigma}(x_1,\ldots,x_n)=x_1w_1+\ldots+x_nw_n-\sigma$ такая, что

$$f(x_1,\,\ldots,x_n)=\begin{cases}1, \text{ если } \sum\limits_{i=1}^{n}x_iw_i-\sigma\geqslant 0;\\ 0, \text{ иначе.}\end{cases}$$

В этом случае говорим, что *линейная форма* $l_{\vec{w},\sigma}$ *задает пороговую функцию* $f(x_1,\ldots,x_n)$, и записывается это так:

$$l_{\vec{w},\sigma}\to f(x_1,\ldots,x_n),$$

или просто $f_{\vec{w},\sigma}$.

Множество всех пороговых функций от n переменных $x_1,\ldots,x_n$ обозначим T^n.

В связи с тем, что линейные формы с целочисленными коэффициентами и порогом позволяют задать любую пороговую функцию, далее в работе рассматриваются только такие линейные формы.

Рассмотрим следующую меру сложности линейной формы

$$L\left(l_{\vec{w},\sigma}\right)=\max\left(|w_1|,\ldots,|w_n|,|\sigma|\right).$$

Рассмотрим следующую меру сложности задания пороговой функции f линейной формой с целочисленными коэффициентами

$$L(f)=\min_{l_{\vec{w},\sigma}\to f}L\left(l_{\vec{w},\sigma}\right),$$

где минимум берется по всем линейным формам с целочисленными коэффициентами и порогом, задающим функцию f.

Рассмотрим две последовательности множеств: $A_1,A_2,\ldots$ и $A'_1,A'_2,\ldots$, где $A_i\supseteq A'_i$ для всех i. Говорят, что «почти все» элементы A_i совпадают с элементами A'_i, если

$$\lim_{i\to\infty}\frac{|A_i|}{|A_i'|}=1.$$

Имеет место следующая теорема, характеризующая величину $L(f)$ для почти всех пороговых функций.

Теорема 1. *Если $n\to\infty$ и $c>1$, то для почти всех пороговых функций $f\in T^n$ выполнено*

$$2^{n-c\cdot\log n}\leqslant L\left(f\right)\leqslant 2^{\frac{1}{2}n\log n+o(n)}.$$

Введем следующую меру сложности линейной формы

$$\mu\left(l_{\vec{w},\sigma}\right)=\sum_{i=1}^{n}|w_i|+|\sigma|\,.$$

Будем называть данную характеристику *весом линейной формы.* Назовем $l_{\vec{w},\sigma}$ минимальной линейной формой, задающей f, если ее вес минимален среди всех линейных форм, задающих f. Весом пороговой функции f назовем вес минимальной линейной формы, задающей данную функцию. Вес пороговой функции обозначим $\mu\left(f\right)$. Имеет место следующее утверждение, характеризующее вес почти всякой пороговой функции.

Теорема 2. *Если $g\left(n\right)\to\infty$ и $g\left(n\right)=o\left(n\right)$, то для почти всех пороговых функций $f\in T^n$ выполнено*

$$2^{n-g(n)}\leqslant\mu\left(f\right)\leqslant 2^{\frac{1}{2}n\log n+o(n)}.$$

Введем понятие расстояния между линейными формами и пороговыми функциями.

Пусть $l_{\vec{w}',\sigma'}$ и $l_{\vec{w}'',\sigma''}$ — линейные формы от n переменных. Расстоянием между линейными формами $l_{\vec{w}',\sigma'}$ и $l_{\vec{w}'',\sigma''}$ назовем следующую величину

$$\rho\left(l_{\vec{w}',\sigma'};l_{\vec{w}'',\sigma''}\right)=\left|\sigma'-\sigma''\right|+\sum_{i=1}^{n}\left|w_i'-w_i''\right|.$$

Эту величину интерпретируем как необходимость сделать ρ последовательных единичных изменений компонент одной линейной формы, чтобы получить другую. *Расстоянием между пороговыми функциями* $f'(x_1,\ldots,x_n)$ и $f''(x_1,\ldots,x_n)$ назовем величину

$$\rho\left(f';f''\right)=\min_{\substack{l_{\vec{w}',\sigma'}\to f'\\ l_{\vec{w}'',\sigma''}\to f''}}\rho\left(l';l''\right).$$

Здесь минимум берется по всем линейным формам $l_{\vec{w}',\sigma'}$ и $l_{\vec{w}'',\sigma''}$, задающим функции f' и f'', соответственно. Сформулируем основной результат данной работы.

Теорема 3. *Если $n\to\infty$, то для почти всех пар пороговых функций $f',f''\in T^n$ выполнено*

$$2^{n+o(n)}\leqslant\rho\left(f',f''\right)\leqslant 2^{\frac{1}{2}n\log n+o(n)}.$$

2. Доказательство теоремы 1

Определим величину $\rho(n)$ следующим образом

$$\rho(n)=\max_{f',f''\in T^n}\rho\left(f';f''\right).$$

Данная величина характеризует расстояние между наиболее удаленными пороговыми функциями от n переменных. Имеет место следующая теорема, доказанная в работе [5].

Теорема 4. *При $n\to\infty$ выполнено*

$$\log_2\rho(n)\sim\frac{1}{2}n\log_2 n.$$

Легко видеть, что $L(n)<\rho(n)$, поэтому верхняя оценка теоремы 1 следует из теоремы 4. Таким образом, для доказательства теоремы 1 достаточно показать нижнюю оценку.

Пусть $\delta=(d_1,\ldots,d_n)$, где $d_i\in\{0,1\}$ для всех $i=1,\ldots,n$. Назовем δ-преобразованием оператор, определенный на множестве T^n и принимающий значения на нем же. Этот оператор записывается так

$$\delta\left[f_{\vec{w},\sigma}\right](x_1,\ldots,x_n)=f_{\vec{w}',\sigma'}(x_1,\ldots,x_n),$$

где

$$\begin{aligned} &w_i' = w_i, \text{ если } d_i = 1,\\ &w_i' = -w_i, \text{ иначе},\\ &\sigma' = \sigma + \sum_{i=1}^{n} (d_i - 1)\, w_i, \end{aligned}$$

при $i = 1, \ldots, n$.

Полученную таким образом пороговую функцию $\delta\left[f_{\vec{w},\sigma}\right]$ будем называть δ-симметричной к f. Понятие δ-симметричности естественным образом обобщается на множества функций: два множества пороговых функций A и B назовем δ-симметричными, если для каждой функции из A найдется δ-симметричная ей функция из B и, наоборот, для каждой функции из B найдется δ-симметричная ей функция из A. Обозначается это так: $B = \delta\,[A]$.

Сигнатурой линейной формы $l_{\vec{w},\sigma} = x_1 w_1 + \ldots + x_n w_n - \sigma$ называем вектор $s\left(l_{\vec{w},\sigma}\right) = (s_1, \ldots, s_n)$, такой что:

$$s_i = \begin{cases} 1, & \text{если } w_i > 0;\\ 0, & \text{иначе}, \end{cases}$$

для $i = 1, \ldots, n$. Имеет место следующая теорема, доказанная в работе [3].

Теорема 5. *Если $l_{\vec{w},\sigma}$ и $l_{\vec{w}',\sigma'}$ задают существенную пороговую функцию f, то $s\left(l_{\vec{w},\sigma}\right) = s\left(l_{\vec{w}',\sigma'}\right)$.*

Из теоремы 5 следует, что сигнатура линейной формы существенной пороговой функции f однозначно определяется этой функцией, что обозначаем $s\,(f)$. Напомним следующий результат работы [3].

Теорема 6 (о сигнатурах). *Отношение равенства сигнатур разбивает множество $\widetilde{T}^n$ на 2^n взаимно симметричных равномощных множества, одним из которых является множество монотонных существенных пороговых функций от n переменных.*

Таким образом, применяя все возможные δ-преобразования к существенной пороговой функции, мы получим 2^n взаимно симметричные пороговые функции. Приведем без доказательства следующий известный результат.

Для T^n — множества всех пороговых функций от n переменных верно утверждение.

Теорема 7. *([2]) При $n \to \infty$ выполнено*

$$|T^n| = 2^{n^2 - n\log n + O(n)}.$$

Доказательство теоремы 1. Очевидно, что существует ровно k^{n+1} наборов $(w_1, \ldots, w_n, \sigma)$, таких что $0 \leqslant w_i < k$ для всех $i = 1, \ldots, n$ и $0 \leqslant \sigma < k$. Обозначим $r(n,k) = k^{n+1}$. Определим наибольшее значение k при фиксированном n, такое что

$$r(n,k) = o(|MT^n|),$$

где MT^n — множество монотонных пороговых функций от n переменных.

Очевидно, что при таком k для почти всех монотонных пороговых функций будет выполнено

$$L(f) > k.$$

По теореме «о сигнатурах» 6 данная оценка будет верна и для произвольной пороговой функции. Оценим мощность множества MT^n. По теореме «о сигнатурах» она может быть оценена следующим образом

$$|MT^n| = \frac{|T^n|}{2^n}.$$

Тогда по теореме 7 имеем

$$|MT^n| = 2^{n^2 - n\log n + O(n)}.$$

Следовательно, задача оценки k может быть сформулирована так: найти наибольшее k, такое что

$$\lim_{n\to\infty} \frac{2^{n^2 - n\log n + O(n)}}{k^{n+1}} = \infty.$$

Упростив выражение, получим

$$\lim_{n\to\infty} 2^{n^2 - n\log n - (n+1)\log k + O(n)} = \infty.$$

$$\lim_{n\to\infty} \left(n^2 - n\log n - (n+1)\log k + O(n)\right) = \infty.$$

Пусть $k = 2^{n-\varepsilon(n)}$, тогда

$$\lim_{n\to\infty}\left(n^2 - n\log n - (n+1)(n-\varepsilon(n)) + O(n)\right) = \infty.$$

$$\lim_{n\to\infty}\left((n+1)\cdot\varepsilon(n) - n\log n + O(n)\right) = \infty.$$

Пусть $\varepsilon(n) = (1+c')\log n$, где $c' > 0$, тогда

$$\lim_{n\to\infty}\left(c'\cdot(n+1)\log n + O(n)\right) = \infty.$$

Это верно при любом $c' > 0$. Следовательно $k \geqslant 2^{n-c\log n}$, где $c > 1$. Теорема доказана.

3. Доказательство теоремы 2

Легко видеть, что $\mu(n) < \rho(n)$, поэтому верхняя оценка теоремы 2 следует из теоремы 4. Таким образом, для доказательства теоремы 2 достаточно показать нижнюю оценку.

Рассмотрим следующее обозначение

$$x^{\underline{m}} = x\cdot(x-1)\cdot\ldots\cdot(x-m+1),$$

где m — целое число. В таком случае говорим, что величина $x^{\underline{m}}$ равна «x в убывающей степени m». Очевидно, что $x^{\underline{m}} = x^m + O\left(x^{m-1}\right)$. Приведем без доказательства два известных результата.

Лемма 1. ([1]) *Если n и m целые, то*

$$\sum_{x=0}^{n-1} x^{\underline{m}} = \frac{1}{m+1}\cdot n^{\underline{m+1}}.$$

Лемма 2. ([1]) *Если n целое, то*

$$x^n = \sum_{k=0}^{n}\left\{\begin{array}{c} n \\ k \end{array}\right\} x^{\underline{k}},$$

где $\left\{\begin{array}{c} n \\ k \end{array}\right\}$ — числа Стирлинга второго рода.

Имеет место следующее утверждение.

Лемма 3. *Если* $p \in \mathbb{N}$, *то*

$$\sum_{x=0}^{n-1} x^m = \frac{1}{m+1} \cdot n^{m+1} + O\left(n^m\right).$$

Доказательство. По лемме 2 выполнено

$$\sum_{x=0}^{n-1} x^m = \sum_{x=0}^{n-1} \sum_{k=0}^{m} \left\{ {m \atop k} \right\} x^{\underline{k}} = \sum_{k=0}^{m} \sum_{x=0}^{n-1} \left\{ {m \atop k} \right\} x^{\underline{k}} = \sum_{k=0}^{m} \left\{ {m \atop k} \right\} \sum_{x=0}^{n-1} x^{\underline{k}}.$$

По лемме 1 получим

$$\sum_{x=0}^{n-1} x^m = \sum_{k=0}^{m} \left\{ {m \atop k} \right\} \cdot \frac{1}{k+1} \cdot n^{\underline{k+1}} =$$

$$= \sum_{k=0}^{m} \left[\left\{ {m \atop k} \right\} \cdot \frac{1}{k+1} \cdot n^{k+1} + O\left(n^k\right) \right] = \left\{ {m \atop m} \right\} \cdot \frac{1}{m+1} \cdot n^{m+1} + O\left(n^m\right).$$

Так как $\left\{ {m \atop m} \right\} = 1$, то

$$\sum_{x=0}^{n-1} x^m = \frac{1}{m+1} \cdot n^{m+1} + O\left(n^m\right).$$

Лемма доказана.

Рассмотрим набор $l = (w_1, \ldots, w_n)$, где $w_i \in \mathbb{N}_0$ для всех $i = 1, \ldots, n$. Рассмотрим множество

$$R_{n,k} = \left\{ l, \sum_{i=1}^{n} w_i < k \right\},$$

где k — натуральное число.

Данное множество состоит из всех наборов l, сумма компонент которых меньше k.

Обозначим $R(n,k)$ мощность множества $R_{n,k}$.

Докажем вспомогательный результат, характеризующий величину $R(n,k)$.

Лемма 4. *При $n \to \infty$ выполнено*

$$R(n,k) = \frac{1}{n!}\left(k^n + O\left(k^{n-1}\right)\right).$$

Доказательство. Легко видеть, что

$$R(2,k) = \frac{k^2+k}{2} = \frac{1}{2}\left(k^2+k\right).$$

Также отметим, что

$$R(n+1,k) = \sum_{i=1}^{k} R(n,i).$$

Следовательно

$$R(3,k) = \frac{1}{2}\cdot\left(\sum_{i=1}^{k} i^2 + \sum_{i=1}^{k} i\right) = \frac{1}{2}\cdot\frac{1}{3}\cdot\left(k^3 + O\left(k^2\right)\right).$$

Таким образом, для того чтобы доказать утверждение леммы достаточно показать, что

$$\sum_{i=1}^{k} i^p = \frac{1}{p+1}\cdot k^{p+1} + O\left(k^p\right).$$

А это верно по лемме 3. Лемма доказана.

Доказательство теоремы 2. Рассмотрим следующую меру сложности линейной формы

$$W\left(l_{\vec{w}}\right) = \sum_{i=1}^{n} |w_i|.$$

Обозначим

$$W(f) = \min_{l_{\vec{w},\sigma}\to f} W\left(l_{\vec{w},\sigma}\right),$$

где минимум берется по всем линейным формам $l_{\vec{w},\sigma}$, задающим f. Очевидно, что $W(f) \leqslant \mu(f)$, поэтому для доказательства теоремы достаточно показать, что

$$W(f) \geqslant 2^{n-g(n)}.$$

Легко видеть, что количество монотонных пороговых функций, для которых $W(f) < k$ не превосходит величину $R(n,k)$. По теореме «о сигнатурах» количество всех пороговых функций, для которых $W(f) < k$ не превосходит величину $2^n \cdot R(n,k)$.

Далее, если при некотором k выполнено

$$2^n \cdot R(n,k) = o(T^n),$$

то можно утверждать, что для почти всех пороговых функций выполнено $W(f) \geqslant k$. Таким образом, необходимо найти наибольшее k, при котором выполнено данное асимптотическое равенство. По лемме 4 имеем

$$R(n,k) = \frac{2^{n\log k} + O\left(k^{n-1}\right)}{2^{n\log n - n\log e + o(n)}} \sim 2^{n\log k - n\log n + n\log e + o(n)}.$$

Теперь найдем наибольшее k, при котором выполнено

$$\lim_{n\to\infty} \frac{2^{n^2 - n\log n + O(n)}}{2^{n\log k - n\log n + n\log e + o(n)}} = \infty.$$

$$\lim_{n\to\infty} \left(n^2 - n\log k + O(n)\right) = \infty.$$

Пусть $k = 2^{n-g(n)}$, тогда

$$\lim_{n\to\infty} (n \cdot g(n) + O(n)) = \infty.$$

Данное выражение верно при всех $g(n)$, таких что $g(n) \to \infty$ и $g(n) = o(n)$. Теорема доказана.

4. Доказательство теоремы 3

Доказажем утверждение теоремы 3.

Верхняя оценка теоремы 3 очевидно следует из теоремы 4. Таким образом, достаточно показать нижнюю оценку.

Заметим, что в связи с утверждением 6 множество линейных форм, задающих пороговые функции, обладает свойством симметрии относительно гиперплоскостей, задаваемых уравнениями вида:

$w_i = 0$, $\sigma = 0$. А следовательно, расстояние ρ между почти всеми пороговыми функциями может быть оценено снизу расстоянием манхеттена между почти всеми точками сферы с центром в нуле и радиусом $\mu(f)$, который имеет место для почти всех пороговых функций f.

Обозначим $S^n(r)$ — манхеттенскую сферу с центром в нуле и радиусом r. Из леммы 4 следует, что если $n \to \infty$, $s(n) = o(r(n))$, то для почти всех наборов $\alpha, \beta \in S^n(r(n))$ выполнено

$$\rho_m(\alpha, \beta) \geqslant s(n).$$

Следовательно, по теореме 2, расстояние ρ между почти всеми пороговыми функциями не менее, чем $2^{n+o(n)}$. Теорема доказана.

Благодарности

Автор благодарит профессора Валерия Борисовича Кудрявцева за постановку задачи, а также профессора Эльяра Эльдаровича Гасанова за ценные обсуждения и советы по работе.

Список литературы

[1] Грэхем Р., Кнут Д., Поташник О. Конкретная математика. М.: Мир, 1998.

[2] Ирматов А. А. Оценки числа пороговых функций // Дискретная математика. 1996. Т. 8, вып. 4. С. 92–107.

[3] Соколов А. П. О конструктивной характеризации пороговых функций // Интеллектуальные системы. 2008. Т. 12, вып. 1–4. С. 363–388.

[4] Соколов А. П. Об одном семействе нейронов с ограниченной сложностью взаимной перестройки // Интеллектуальные системы. 2009. Т. 13, вып. 1–4. С. 477–490.

[5] Соколов А. П. Асимптотика логарифма сложности перестройки нейронов // Интеллектуальные системы. 2009. Т. 13, вып. 1–4. С. 119–128.