

Коцовський В. М.

кандидат технічних наук

ДВНЗ "Ужгородський національний університет"

Кількісні оцінки розпізнавальної здатності двопорогових нейронних елементів

Робота присвячена вивченню властивостей двопорогових нейронних елементів, які є одним з найпростіших узагальнень класичних нейроелементів МакКаллока-Піттса. Використання двопорогових нейронів дає змогу подолати деякі обмеження, притаманні звичайним пороговим елементам, зокрема знайти розв'язок відомої XOR-проблеми. У роботі вивчені питання, які стосуються оцінки кількості дихотомій скінченної множини у n -вимірному просторі, які можна отримати за допомогою двопорогових нейронів. Також досліджено асимптотичну поведінку кількості дихотомій та розглянуто питання знаходження розмірності Вапніка-Червоненкіса двопорогових нейроелементів.

Ключові слова: нейронний елемент, двопороговий нейрон, штучна нейромережа, розпізнавання.

Количественные оценки распознавательной мощности двупороговых нейронных элементов

Робота посвящена изучению свойств двупороговых нейронных элементов, которые являются одним из самых простых обобщений классических нейроэлементов МакКаллока-Питтса. Использование двупороговых нейронов дает возможность решить известную XOR-проблему. В работе изучены вопросы, касающиеся оценки числа дихотомий конечного множества в n -мерном пространстве, которые можно получить с помощью двупороговых нейронов. Исследовано асимптотическое поведение этого числа и рассмотрен связанный с этим вопрос оценки размерности Вапника-Червоненкиса двупороговых нейронов.

Ключевые слова: нейронный элемент, двупороговий нейрон, искусственная нейросеть, распознавание.

Quantitative estimations of the capability of bithreshold neural units

The given paper is devoted to the study of the properties of the simplest multithreshold generalization of McCulloch-Pitts neurons, namely bithreshold neurons with linear input operator. Usage of neuron supplied with two thresholds provides the possibility to find out the solution of the famous XOR-problem. The most frequent quantitative characteristic of representative power of neuron-like units with discrete activation function is the number of all possible dichotomies of the finite subset of n -dimensional space achieved by using such devices. The asymptotic behavior of this number is given in the paper. The related question of Vapnik-Chervonenkis dimension of bithreshold neuron is also studied.

Key words: neural unit, bithreshold neuron, artificial neural network, recognition.

1. Вступ

Конекціоністський підхід заснований на використанні нейромереж до моделювання складних об'єктів та явищ. Кілька останніх десятиріч штучні

нейромережі та інші нейроподібні структури широко використовуються для розв'язування широкого кола актуальних господарських задач [1]. Однією з ключових проблем, які постають при практичному застосуванні штучних нейронних мереж є проблема вибору моделі нейронів, які утворюють нейромережу. Традиційними вважаються підходи з використання нейронів з лінійним вхідним оператором та функцію активації порогового типу (пороговий елемент) або сигмоїдального типу (неперервний нейрон). Кожний з цих підходів має свої переваги і недоліки [2].

Уведення у розгляд двопорогових нейронних елементів (ДНЕ) та їх дослідження мотивується у літературі більш потужними можливостями цих елементів порівняно із звичайними пороговими елементами по розпізнаванню належності точок у \mathbb{R}^n до одного з двох заданих класів. Це саме твердження стосується і нейронів із неперервними функціями активації двопорогового типу, до яких належить широкий клас "дзвіноподібних" функцій.

У роботах [3, 4] зроблено спробу кількісно оцінити переваги використання двопорогових нейроелементів. Зокрема, авторами були встановлені оцінки кількості класифікацій точок скінченної множини у дійсному векторному просторі, які можна отримати за допомогою ДНЕ. Слід зауважити, що у асимптотичному сенсі встановленні оцінки мають різний порядок росту. У даній роботі з використанням прийомів теорії лінійних нерівностей вдалося отримати та обґрунтувати уточненні оцінки кількості різних двопорогових класифікацій.

2. Модель двопорогового нейрона

Нехай \mathbb{R}^n — n -вимірний дійсний евклідов простір. Якщо $\mathbf{w} = (w_1, w_2, \dots, w_n) \in \mathbb{R}^n$, $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, то величину скалярного добутку

$$(\mathbf{w}, \mathbf{x}) = \sum_{i=1}^n w_i x_i$$

будемо називати зваженою сумою, що відповідає вектору \mathbf{x} .

Штучним двопороговим дійсним нейронним елементом з ваговим вектором $\mathbf{w} \in \mathbb{R}^n$, порогами $t_1, t_2 \in \mathbb{R}$ ($t_1 < t_2$) будемо називати функціональний

елемент з n дійсними входами x_1, x_2, \dots, x_n та одним виходом $y \in \{-1, 1\}$, поведінка якого описується співвідношеннями:

$$y = \begin{cases} -1, & t_1 < (\mathbf{w}, \mathbf{x}) < t_2, \\ 1, & (\mathbf{w}, \mathbf{x}) \leq t_1 \vee (\mathbf{w}, \mathbf{x}) \geq t_2. \end{cases}$$

В термінах [2] зважена сума є вхідним оператором ДНЕ із функцією активації вигляду

$$f_{t_1, t_2}(x) = \begin{cases} -1, & t_1 < x < t_2, \\ 1, & x < t_1 \vee x > t_2. \end{cases}$$

ДНЕ повністю визначається впорядкованою трійкою (\mathbf{w}, t_1, t_2) , яку будемо надалі називати вектором структури або просто структурою ДНЕ.

ДНЕ із структурою (\mathbf{w}, t_1, t_2) здійснює розбиття простору \mathbb{R}^n на дві підмножини наступним чином:

$$\mathbb{R}^{n+} = \{\mathbf{x} \in \mathbb{R}^n \mid (\mathbf{w}, \mathbf{x}) \leq t_1\} \cup \{\mathbf{x} \in \mathbb{R}^n \mid (\mathbf{w}, \mathbf{x}) \geq t_2\}, \quad \mathbb{R}^{n-} = \{\mathbf{x} \in \mathbb{R}^n \mid t_1 < (\mathbf{w}, \mathbf{x}) < t_2\}.$$

Дві множини $A^+ \subset \mathbb{R}^n$ і $A^- \subset \mathbb{R}^n$ назвемо двопорогово-сепарабельними (д-сепарабельними), якщо знайдеться такий ДНЕ із вектором структури (\mathbf{w}, t_1, t_2) , що $A^+ \subset \mathbb{R}^{n+}$ і $A^- \subset \mathbb{R}^{n-}$. У цьому випадку будемо казати, що ДНЕ із структурою (\mathbf{w}, t_1, t_2) здійснює д-розбиття (A^+, A^-) множини $A = A^+ \cup A^-$ на дві множини A^+ і A^- , що не перетинаються. Якщо ДНЕ має структуру (\mathbf{w}, t_1, t_2) , то для довільної множини $A \subset \mathbb{R}^n$ її підмножини $A^+ = A \cap \mathbb{R}^{n+}$, $A^- = A \cap \mathbb{R}^{n-}$ є д-сепарабельними. Легко бачити, що д-сепарабельність множин є узагальненням лінійної сепарабельності. Вже у одновимірному просторі можна навести приклад д-сепарабельних множин, які не є лінійно сепарабельними (наприклад $A^- = \{0\}$ $A^+ = \{-1, 1\}$).

3. Розпізнавальна здатність ДНЕ

Нехай $D_1(A)$ — кількість способів розбиття точок скінченної множини A на два класи за допомогою звичайного одношарового перцептрона (порогового

елемента). Надалі будемо вважати, що $A \subset \mathbb{R}^n$ і $\text{Card } A = m$. Нехай $D_1(m, n) = \max \{D_1(A) \mid A \subset \mathbb{R}^n, \text{Card } A = m\}$. Добре відомо [1, 3], що

$$D_1(m, n) = \begin{cases} 2 \sum_{i=0}^n C_{m-1}^i, & m > n + 1, \\ 2^m, & m \leq n + 1, \end{cases} \quad (1)$$

де C_{m-1}^i — біномні коефіцієнти, причому $D_1(A) = D_1(m, n)$ тоді і тільки тоді, коли усі точки множини A знаходяться у загальному положенні (через жодні $n + 1$ точки множини A не можна провести гіперплощину).

Нехай $D_2(A)$ — кількість різних д-розбиттів (A^+, A^-) скінченної m -елементної множини $A \subset \mathbb{R}^n$, $D_2(m, n) = \max \{D_2(A) \mid A \subset \mathbb{R}^n, \text{Card } A = m\}$. Покажемо, що (1) можна використати для того, щоб отримати оцінки для $D_2(m, n)$, причому верхня оцінка значно покращує оцінку, отриману у роботах [3, 4]. Надалі без додаткових застережень будемо вважати, що $A \subset \mathbb{R}^n$ і $\text{Card } A = m$.

Твердження 1. *При $m > n$ має місце нерівність*

$$D_2(m, n) < 2 \sum_{i=0}^{n+1} C_{2m-1}^i. \quad (2)$$

Якщо елементи множини A знаходяться у загальному положенні і $m > n + 1$, то

$$D_2(A) \geq 2 \sum_{i=0}^{n+1} C_{m-1}^i - 1. \quad (3)$$

Доведення. Для кожного вектора $\mathbf{x} = (x_1, \dots, x_n) \in A$ побудуємо два $n + 2$ -вимірні вектори $\mathbf{x}' = (-x_1, \dots, -x_n, 1, 0)$ і $\mathbf{x}'' = (x_1, \dots, x_n, 0, -1)$. Якщо ДНЕ із структурою (\mathbf{w}, t_1, t_2) здійснює д-розбиття множини A , то гіперплощина $w_1 x_1 + \dots + w_n x_n + t_1 x_{n+1} + t_2 x_{n+2} = 0$ здійснює лінійне розбиття (B^+, B^-) множини

$B = \{\mathbf{x}' | \mathbf{x} \in A\} \cup \{\mathbf{x}'' | \mathbf{x} \in A\}$ і $B_1^- = \{\mathbf{x}' | \varphi_{\mathbf{w}, t_1, t_2}(\mathbf{x}) < 0\} \cup \{\mathbf{x}'' | \varphi_{\mathbf{w}, t_1, t_2}(\mathbf{x}) < 0\} \subset B^-$,
 $B_1^+ = \{\mathbf{x}' | \mathbf{x} \in A, (\mathbf{w}, \mathbf{x}) \leq t_1\} \cup \{\mathbf{x}'' | \mathbf{x} \in A, (\mathbf{w}, \mathbf{x}) \geq t_2\} \subset B^+$. Тому кількість різних д-розбиттів множини A не перевищує кількість лінійних однорідних розбиттів множини B (насправді має місце строга нерівність, бо B_1^+ — власна підмножина множини B^+). Відомо [1], що кількість лінійних однорідних розбиттів множини можна обчислити за формулою (1), зменшивши на 1 кількість доданків. Звідси випливає справедливість нерівності (2).

Оцінимо порядок росту правої частини у (2). Відомо [5], що при $l > n + 1$

$$\sum_{i=0}^n C_{l-1}^i \leq 1,5 \frac{l^n}{n!}.$$

З урахуванням останньої нерівності маємо, що

$$D_2(m, n) < 3 \cdot \frac{(2m)^{n+1}}{(n+1)!} = O(m^{n+1}). \quad (4)$$

Для порівняння можна нагадати, що $D_1(m, n) = O(m^n)$ [1, 5]. Оцінка (4) є значно кращою, ніж верхня оцінка, наведена у роботі [3], де фактично показано, що $D_2(m, n) < O(m^{2n+1})$.

Перейдемо до встановлення нижньої оцінки. Якщо при встановленні верхньої оцінки нами фактично було здійснено перехід від простору \mathbb{R}^n до простору \mathbb{R}^{n+2} , то для з'ясування нижньої оцінки перейдемо від $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ до $\tilde{\mathbf{x}} = (x_1, \dots, x_n, -1) \in \mathbb{R}^{n+1}$. Нехай $\tilde{\mathbf{w}} = (w_1, \dots, w_n, t_1)$. Тоді

$$\forall \mathbf{x} \in A \quad \mathbf{x} \in A^- \Leftrightarrow 0 < (\tilde{\mathbf{w}}, \tilde{\mathbf{x}}) < t_2 - t_1.$$

Очевидно, що якщо ДНЕ із структурою (\mathbf{w}, t_1, t_2) здійснює д-розбиття множини A , то гіперплощина $w_1 x_1 + \dots + w_n x_n + t_1 x_{n+1} = t_2 - t_1$ розбиває множину

$C = \{\tilde{\mathbf{x}} | \mathbf{x} \in A\}$ на дві множини $C^+ = \{\tilde{\mathbf{x}} | (\tilde{\mathbf{w}}, \tilde{\mathbf{x}}) \geq t_2 - t_1\}$ і $C^- = C \setminus C^+$. Тоді з урахуванням (1) отримуємо (3). Твердження доведене.

Розглянемо останній доданок у (3). Легко бачити, що при фіксованому n він є величиною порядку $\theta(m^{n+1})$. Тому нижня оцінка має той самий асимптотичний порядок росту, що й верхня оцінка (4).

Зауваження. Нижня оцінка (3) за допомогою інших методів була встановлена у роботах [3, 4] для двопорогових елементів більш загального вигляду.

Покажемо, що верхня оцінка (2) для $D_2(m, n)$ може бути використана для доведення наступного факту

Твердження 2. При $n > 10$ для довільної множини $A \subset \mathbb{R}^n$, такої, що $\text{Card } A \geq 5n$ знайдеться розбиття множини A , яке не є d -розбиттям.

Для доведення використаємо (4) та формулу Стірлінга. Отримаємо

$$D_2(5n, n) < 3 \cdot \frac{(10n)^{n+1}}{(n+1)!} = \frac{30 \cdot 10^n n^{n+1}}{(n+1)!} = \frac{30 \cdot 10^n n^{n+1} e^{n+1}}{\sqrt{2\pi(n+1)}(n+1)^{n+1} e^{\theta(n+1)}} < 6 \cdot (10e)^n,$$

де $|\theta(n)| < (12n)^{-1}$. Якщо

$$n > \frac{\ln 6}{4 \ln 2 - \ln 5e} \approx 10,9822,$$

то $D(5n, n) < 2^{5n}$. Ми отримали, що кількість d -розбиттів множини A менша за кількість усіх її підмножин. Звідси випливає існування розбиттів множини A , які не є d -розбиттями.

Встановимо тепер обмеження на потужність множини A , яке б забезпечувало d -сепарабельність довільного розбиття множини A , тобто виконання умови $D_2(m, n) = 2^m$.

Твердження 3. Для всіх $n \in \mathbb{N}$ можна вказати таку множину $A_n \subset \mathbb{R}^n$, що $\text{Card } A_n = 2n$ і $D_2(A_n) = 2^{2n}$.

Для доведення використаємо індукцію по розмірності векторного простору, причому будемо доводити, що усі розбиття множин $n \in \mathbb{N}$ можна здійснити на ДНЕ, пороги яких задовольняють умову

$$t_1 = -1, \quad t_2 = 2. \quad (5)$$

Справедливість твердження при $n=1$ для довільної множини, яка містить дві точки дійсної прямої легко отримати безпосередньою перевіркою (усі розбиття двохелементної множини є лінійно сепарабельними). Крім того легко переконатися, що усі д-розбиття при $n=1$ можна отримати за допомогою з порогамі, які задовольняють умову (5). Для визначеності покладемо $A_1 = \{-1, 1\}$. Нехай для всіх $r < n$ твердження вже доведене. Доведемо його при $r=n$. Задамо множину $A_n \subset \mathbb{R}^n$ наступним чином

$$A^n = \{(x_1, \dots, x_{n-1}, 0) \mid (x_1, \dots, x_{n-1}) \in A^{n-1}\} \cup \{(0, \dots, 0, -1), (0, \dots, 0, 1)\}.$$

Покажемо, що усі розбиття множини A_n є д-розбиттями. Нехай (A_n^+, A_n^-) — довільне розбиття множини A_n і нехай (A_{n-1}^+, A_{n-1}^-) — відповідне розбиття множини A_{n-1} ($A_{n-1}^+ = \{(x_1, \dots, x_{n-1}) \mid (x_1, \dots, x_{n-1}, 0) \in A_n^+\}$, $A_{n-1}^- = A_{n-1} \setminus A_{n-1}^+$). За припущенням індукції його можна здійснити за допомогою ДНЕ із структурою $((w_1, \dots, w_{n-1}), -1, 2)$. Можливими є 4 випадки:

1. $\{(0, \dots, 0, -1), (0, \dots, 0, 1)\} \subset A_n^+$. У цьому випадку покладемо $w_n = 3$;
2. $(0, \dots, 0, -1) \in A_n^+$, $(0, \dots, 0, 1) \in A_n^-$. У цьому випадку покладемо $w_n = 1,5$;
3. $(0, \dots, 0, -1) \in A_n^-$, $(0, \dots, 0, 1) \in A_n^+$. У цьому випадку покладемо $w_n = -1,5$;
4. $\{(0, \dots, 0, -1), (0, \dots, 0, 1)\} \subset A_n^-$. У цьому випадку покладемо $w_n = 0$.

Легко переконатися в тому, що у кожного з чотирьох випадків ДНЕ із структурою $((w_1, \dots, w_{n-1}, w_n), -1, 2)$ здійснює д-розбиття множини A_n , причому пороги t_1, t_2 задовольняють (5). Твердження доведене.

Наслідок. Якщо LBT_n — множина усіх n -місних ДНЕ, то при $n > 10$

$$2n \leq \text{VCDim}(LBT_n, \mathbb{R}^n) < 5n,$$

де $\text{VCDim}(LBT_n, \mathbb{R}^n)$ — розмірність Вапніка-Червоненкіса.

4. Висновки

У роботі розглянуто питання оцінки потенційної спроможності двопо-
рогових нейронних елементів вирішувати задачі розпізнавання підмножин n -
вимірного евклідового простору. Встановлено, що за допомогою ДНЕ можна
правильно розпізнати $\theta(m^{n+1})$ різних дихотомій, де m — потужність множини,
елементи якої знаходяться у загальному положенні. Отриманий результат є по-
кращенням оцінок, наведених у [3, 4], і дозволяє кількісно оцінити потенційні
переваги від застосування у розпізнавальних пристроях ДНЕ замість класичних
порогових елементів, для яких кількість відповідних дихотомій є величиною,
порядок росту якої не перевищує $O(m^n)$.

Також знайдено оцінки розмірності Вапніка-Червоненкіса для класу ДНЕ
із n входами — ключового параметру теорії навчання.

Список літератури

1. Хайкин, С. Нейронные сети: полный курс / С. Хайкин. – 2-е изд. – М.: Вильямс-Телеком, 2006. – 1104 с.
2. Руденко, О. Г. Штучні нейронні мережі / О. Г. Руденко, Є. В. Бодянский. – Харків: ТОВ «Компанія СМІТ», 2006. – 404 с.
3. Olafsson, S. The capacity of multilevel threshold function / S. Olafsson and Y. S. Abu-Mostafa // IEEE Trans. Pattern Anal. Machine Intell. – 1988, vol. 10, NO. 2. – PP. 277-281.
4. Takiyama, R. Multiple threshold perceptron / R. Takiyama // Pattern recognition. – 1978, vol. 10. – PP. 27-30.
5. Вапник, В. Н. Теория распознавания образов. Статистические проблемы обучения / В. Н. Вапник, А. Я. Червоненкис. – М.: Наука, 1974. – 416 с.