УДК 681.14

Коцовський[*] В.М., асистент УжНУ

Нейронні елементи над полем комплексних чисел

В роботі розглядаються нейронні елементи з комплексними вагами. Наведені алгоритми навчання цих нейроннів для деяких видів функцій активації.

Ключові слова: нейронний елемент, нейромережа, порогова логіка.

[*]*E-mail:* cyber@mail.uzhgorod.ua

V. M. Kotsovsky, assistant UzhNU

Neural elements over the complex field

Neural elements with complex weights have been studied in the paper. Learning algorithms for some kind of activation functions are given in the paper.

Key Words: neural element, neural network, threshold logic.

Нейронні елементи (НЕ) використовуються при побудові нейромереж, які є ефективним механізмом вирішення широкого кола задач класифікації об'єктів, розпізнавання образів, стиску інформації, прогнозування поведінки динамічних систем, наближення і екстраполяції функцій. При цьому застосовуються різноманітні підходи по визначенню конфігурації НЕ і зв'язків між ними у нейромережах. В роботах [1-7] розглядаються НЕ як з неперервними, так і з розривними функціями активації. Перші з них використовуються для наближення неперервних функцій, другі – для наближення (здебільшого точної реалізації) дискретних функцій. Особливо важливим є вивчення властивостей булювих функцій, які можна задати на НЕ за допомогою різних функцій активації.

Розглянемо класичний підхід щодо реалізації булювих функцій на НЕ. Розглядаються булеві функції в алфавітах $Z_2=\{0,1\}$ або $H_2=\{-1,1\}$. Якщо для булевої функції $f:H_2^n\to H_2$ знайдеться такий n+1-вимірний дійсний вектор $w=(w_1,\dots,w_n,w_{n+1})\in R^{n+1}$, що для всіх векторів $(x_1,\dots,x_n)\in Z_2^n$ має місце співвідношення $\sum_{j=1}^{n}w_jx_j+w_{n+1}<0\Leftrightarrow f(x_1,\dots,x_n)=-1$, то кажуть, що булева функція $f(x_1,\dots,x_n)$ є нейрофункцією (пороговою функцією) і реалізується на НЕ з ваговим вектором $\boldsymbol{w}\in R^{n+1}$. Скалярний добуток $(w,x)= =\sum_{j=1}^{n}w_jx_j+w_{n+1}$, де $x=(x_1,\dots x_n,1)$ називають зваженою сумою, що відповідає вхідному вектору *x*. Надалі будемо використовувати позначення $f(x)=f(x_1,\dots,x_n)$, вважаючи, що функція *f*(*x*) не залежить від змінної x_{n+1}.

З а у в а ж е н н я 1. В деяких випадках зручно вважати, що x_{n+1} дорівнює деякій наперед визначеній константі, яка відмінна від нуля.

У зв'язку з "невеликою" потужністю класу булевих нейрофункцій (див. [3]) багатьма авторами робилися спроби розширити його за рахунок НЕ більш загального вигляду. Одним з можливих підходів є розгляд комплексних булевих нейрофункцій в алфавіті $\{\alpha,\beta\}\subset C$, які породжуються ваговим вектором $w\in C^{n+1}$, де C – поле комплексних чисел. Будемо визначати знакову функцію $\mathrm{Lsign}:C\to H_2$ так: нехай l – довільна пряма на площині, яка розбиває її на дві півплощини C_- і C_+. Тоді

$$\mathrm{Lsign}\, z=\begin{cases}-1, & z\in C_-,\\ 1, & z\in C_+\cup l.\end{cases}$$

Застосовуючи до зважених суми $(w,\bar{z})$ знакову функцію Lsign, отримаємо відображення $f:\{\alpha,\beta\}^n\to H_2$, яке будемо називати комплексною булевою нейрофункцією в алфавіті $\{\alpha,\beta\}$. Визначення знакової функції Lsign є досить загальним. Покажемо, що його можна спростити, не звузивши при цьому класу комплексних булевих нейрофункцій. Справді, за рахунок вибору вільного члена w_{n+1} у зваженій сумі $(w,\bar{z})$, домножання всіх вагових коефіцієнтів на множник вигляду $e^{i\varphi}$ і певних змін вагового вектора [1], без втрати загальності можна обмежитися наступною знаковою функцією:

$$\operatorname{Resign} z = \begin{cases} -1, & \operatorname{Re} z < 0, \\ 1, & \operatorname{Re} z > 0. \end{cases}$$

Схожий підхід до визначення знакових функцій був застосований в [1]. Позначимо через $P(\{\alpha,\beta\},C)$ клас всіх комплексних нейрофункцій в алфавіті $\{\alpha,\beta\}$. Виникає питання, як пов'язані між собою класи комплексних нейрофункцій у різних алфавітах. Відповідь на це питання дає

Лема 1. Для довільних алфавітів $\{\alpha,\beta\}$, $\{\gamma,\delta\}$ має місце взаємно однозначна відповідність $P(\{\alpha,\beta\},C) \leftrightarrow P(\{\gamma,\delta\},C)$.

Доведення. Нехай $f(z) \in P(\{\alpha,\beta\},C)$. Тоді знайдеться такий вектор $w \in C^{n+1}$, що для всіх $z \in \{\alpha,\beta\}^n$ $f(z) = \operatorname{Resign}(w,\bar{z})$. Відображення $z' \to \frac{\beta-\alpha}{\delta-\gamma}(z'-\gamma)+\alpha$ здійснює взаємно однозначну відповідність між елементами множин $\{\gamma,\delta\}^n$ і $\{\alpha,\beta\}^n$. Тоді

$$(w,\bar{z}) = \sum_{j=1}^{n} w_j z_j + w_{n+1} = \frac{\beta-\alpha}{\delta-\gamma}\sum_{j=1}^{n} w_j z'_j + \left(\alpha - \frac{\beta-\alpha}{\delta-\gamma}\gamma\right)\sum_{j=1}^{n} w_j z'_j + w_{n+1} = (w',\overline{z'}),$$

де

$$w'_j = \frac{\beta-\alpha}{\delta-\gamma} w_j,\ (j=1,2,\ldots n),$$

$$w'_{n+1} = \left(\alpha - \frac{\beta-\alpha}{\delta-\gamma}\gamma\right)\sum_{j=1}^{n} w_j + w_{n+1}.$$

Нехай $g(z')$ – бульова функція в алфавіті $\{\gamma,\delta\}$, яка реалізується на НЕ з ваговим вектором w'. Легко бачити, що відображення $f \to g$ є бієкцією між елементами множин $P(\{\alpha,\beta\},C)$ і $P(\{\gamma,\delta\},C)$. Лема доведена.

Зауваження 2. Лема 1 показує, що не можна за рахунок вибору алфавіту отримати клас бульових функцій, який би мав більшу потужність, ніж клас $P(H_2,C)$.

Виникає питання про зв'язок між класами $P(\{\alpha,\beta\},C)$ і $P(\{\alpha,\beta\},D)$, де D – деяка власна підмножина поля C.

Лема 2. Якщо для алфавіту $\{\alpha,\beta\}$ виконується умова $\operatorname{Re}\alpha \neq \operatorname{Re}\beta$, то

$$P(\{\alpha,\beta\},C) = P(\{\alpha,\beta\},R).$$

Доведення. Нехай $P_n(\{\alpha,\beta\},C)$ – множина всіх n-місних комплексних нейрофункцій в алфавіті $\{\alpha,\beta\}$. Покажемо, що для всіх натуральних n $P_n(\{\alpha,\beta\},C) = P_n(\{\alpha,\beta\},R)$. Згідно леми 1 $P_n(\{\alpha,\beta\},C) \leftrightarrow P_n(\{\operatorname{Re}\alpha,\operatorname{Re}\beta\},C)$. Нехай f – довільний представник класу $P_n(\{\operatorname{Re}\alpha,\operatorname{Re}\beta\},C)$, $z \in \{\alpha,\beta\}^n \times \{1\}, z_j = x_j + iy_j, x_j, y_j \in R, (j=1,\ldots,n)$, $w_j = u_j + iv_j$, $u_j, v_j \in R$, $(j=1,\ldots,n+1)$. Тоді

$$f(z) = \operatorname{Re}\left(\sum_{j=1}^{n} w_j x_j + w_{n+1}\right) = \sum_{j=1}^{n} u_j x_j + u_{n+1} = \operatorname{Re}\left(\sum_{j=1}^{n} u_j z_j + u_{n+1}\right).$$

З останньої рівності випливає, що класи бульових функцій $P_n(\{\operatorname{Re}\alpha,\operatorname{Re}\beta\},C)$ і $P_n(\{\alpha,\beta\},R)$ мають однакову потужність. Тоді множини $P_n(\{\alpha,\beta\},R)$ і $P_n(\{\alpha,\beta\},C)$ також мають однакову потужність. Оскільки перша з них є підмножиною другої, то вони є рівними. Лема доведена.

Зауваження 3. Для алфавіту H_2 лема 2 доведена в роботі [2].

З доведеної леми випливає, що НЕ, коефіцієнти вагових векторів яких належать дійсній осі, породжують клас усіх комплексних нейрофункцій. Виявляється, що цей факт є типовим і має місце для НЕ, вагові коефіцієнти яких належать деякій прямій у комплексній площині.

Теорема 1. Якщо $R(\gamma) = \{\gamma x \mid x \in R^{n+1}\}$ і для комплексних чисел α,β,γ виконуються умови $|\operatorname{Arg}\gamma| < \frac{\pi}{2}$, $\operatorname{Re}(\alpha-\beta)\gamma \neq 0$, то класи бульових нейрофункцій $P(\{\alpha,\beta\},C)$ і $P(\{\alpha,\beta\},R(\gamma))$ співпадають.

Доведення. Розглянемо довільну нейрофункцію $f(z) \in P(\{\alpha,\beta\},C)$. Тоді знайдеться такий вектор $w \in C^{n+1}$, що для всіх $z \in \{\alpha,\beta\}^n$ має місце рівність $\operatorname{Resign}(w,\bar{z}) = f(z)$. Звідси

$$(w,\bar{z}) = \sum_{j=1}^{n} w_j z_j + w_{n+1} = \sum_{j=1}^{n} w_j \gamma^{-1}\gamma z_j + w_{n+1} = (w',\overline{z'})$$

, де $w'_j = w_j\gamma^{-1}$, $z'_j = \gamma z_j$ $(j=1,\ldots,n)$ $w'_{n+1} = w_{n+1}$. Тому кожній комплексній нейрофункції $f(z)$ в алфавіті $\{\alpha,\beta\}$ взаємно однозначно відповідає нейрофункція $g(z')$ в алфавіті $\{\gamma\alpha,\gamma\beta\}$, причому для всіх $z \in \{\alpha,\beta\}^n$ має місце рівність $f(z) = g(z')$. Застосувавши лему 2 до функції $g(z')$, отримаємо

$$\operatorname{Re}(w, \bar{z}) = \operatorname{Re}(w', \overline{z'}) = \operatorname{Re}\left(\sum_{j=1}^{n} u_j z'_j + u_{n+1}\right) =$$

$$\operatorname{Re}\left(\sum_{j=1}^{n} u_j \gamma \cdot z_j + u_{n+1}\right) = \operatorname{Re}(\tilde{w}, \bar{z}),$$

де $u_j \in R,\ \tilde{w}_j = u_j \gamma,\ (j = 1,\ldots,n),\ \tilde{w}_{n+1} = \dfrac{u_{n+1}\gamma}{\operatorname{Re}\gamma}$.

Отже, бульова функція $f(z)$ реалізується на НЕ з ваговим вектором $\tilde{w} \in R(\gamma)$. Теорема доведена.

Виникає питання про алгоритм знаходження для комплексної нейрофункції вагових коефіцієнтів НЕ, який реалізує цю функцію, тобто про алгоритм навчання НЕ, при умові, що $w \in R(\gamma)$.

Нехай A_+, A_- – скінченні підмножини, які складаються з векторів, що належать множині $C^n \times \{\gamma\}$, такі, що $A_+ \cap A_- = \varnothing$. Множини A_+ і A_- назвемо γ-відділимими, якщо знайдеться такий вектор $w \in R(\gamma)$, що для всіх $z \in A = A_+ \cup A_-$ мають місце співвідношення

$$(w, \bar{z}) > 0, \quad \text{якщо } z \in A_+,$$
$$(w, \bar{z}) < 0, \quad \text{якщо } z \in A_-.$$

Крім того, будемо вважати, що

$$\forall z \in A \ \left|\operatorname{Re}\left(e^{i\varphi} z_j\right)\right| \ge c > 0 \ \ (j = 1,\ldots n), \qquad (1)$$

Нехай послідовність векторів $\{z^k\}$ (навчальна послідовність) задовольняє наступні дві умови: 1) $z^k \in A,\ k \in N$; 2) кожний елемент множини A входить у цю послідовність нескінченну кількість раз. Не втрачаючи загальності, будемо вважати, що $\gamma = e^{i\varphi}$, де $-\dfrac{\pi}{2} < \varphi < \dfrac{\pi}{2}$. Покладемо $w^0 = (0,\ldots,0)$ і будемо будувати послідовність векторів $\{w^k\}$ наступним чином:

$$w^k = w^{k-1} + t_k h_\varphi\left(z^k\right) e^{i\varphi}, \qquad (2)$$

де $h_\varphi(z) = \left(\operatorname{Re}\left(\overline{z_1} e^{-i\varphi}\right),\ldots,\operatorname{Re}\left(\overline{z_n} e^{-i\varphi}\right),1\right)$, а коефіцієнт t_k визначається так:

$$t_k = \begin{cases} 1, & \text{якщо } \operatorname{Re}\left(w^{k-1}, \overline{z^k}\right) \le 0 \text{ і } z \in A_+, \\ -1, & \text{якщо } \operatorname{Re}\left(w^{k-1}, \overline{z^k}\right) \ge 0 \text{ і } z \in A_-, \\ 0, & \text{в інших випадках.} \end{cases} \qquad (3)$$

Алгоритм побудови послідовності $\{w^k\}$ за правилом (2) назвемо алгоритмом навчання НЕ.

Теорема 2. Якщо скінченні множини A_+ і A_- є γ-відділимими, то послідовність вагових векторів (2) стабілізується через деяку кількість кроків на деякому векторі w^m, який відокремлює ці множини.

Доведення. Припустимо противне. Будемо вважати, що на кожному кроці алгоритму $t_k \ne 0$ (можна відкинути такі z^k, для яких $t_k = 0$, оскільки на відповідних їм кроках алгоритму не змінюються вагові вектори). Тоді $w^{m+1} = t_1 h_\varphi\left(z^1\right) e^{i\varphi} + \ldots + t_m h_\varphi\left(z^m\right) e^{i\varphi}$. Помножимо останню рівність скалярно на вектор $w \in R(\gamma)$, який відокремлює множини A_+ і A_-. Без обмеження загальності можна вважати, що знайдеться таке $d > 0$, що $\forall z \in A$ виконується нерівність $\left|\left(w, h_\varphi(z)\right)\right| \ge d > 0$ (цього можна досягти змінивши відповідним чином вільний член w_{n+1}). Тоді з нерівності Коші-Шварца отримаємо, що

$$\|w\| \cdot \left\|w^{m+1}\right\| \ge \left|\left(w, w^{m+1}\right)\right| \ge \sum_{k=1}^{m} \left|\left(w, h_\varphi\left(z^k\right)\right)\right| \ge md.$$

З останньої нерівності випливає, що

$$\left\|w^{m+1}\right\|^2 \ge \frac{m^2 d^2}{\|w\|^2}. \qquad (4)$$

З іншого боку, підносячи рівність (2) скалярно до квадрату, отримаємо:

$$\left\|w^{k+1}\right\| = \left\|w^k\right\| + 2t_{k+1} \operatorname{Re}\left(w^k, h_\varphi\left(z^{k+1}\right) e^{i\varphi}\right) + \left\|h_\varphi\left(z^{k+1}\right)\right\|^2.$$

Відповідно до алгоритму навчання всі вектори w^k задовольняють умову $w^k = u^k e^{i\varphi}$, де $u^k \in R^{n+1}$. Тому

$$\operatorname{Re}\left(w^k, h_\varphi\left(z^{k+1}\right) e^{i\varphi}\right) = \operatorname{Re} \sum_{j=1}^{n} u_j^k e^{i\varphi} \operatorname{Re}\left(\overline{z_j^{k+1}} e^{-i\varphi}\right) \cdot \overline{e^{i\varphi}} +$$

$$+ \operatorname{Re}\left(u_{n+1}^k e^{i\varphi} \overline{e^{i\varphi}}\right) = \operatorname{Re} \sum_{j=1}^{n} u_j^k \left(x_j^{k+1} \cos\varphi - y_j^{k+1} \sin\varphi\right) +$$

$$+ u_{n+1}^k = \operatorname{Re}\left(\sum_{j=1}^{n} \operatorname{Re}\left(u_j^k e^{i\varphi}\left(x_j^{k+1} + i y_j^{k+1}\right)\right) + u_{n+1}^k e^{i\varphi} \overline{e^{i\varphi}}\right) =$$

$$= \operatorname{Re}\left(\sum_{j=1}^{n} w_j^k z_j^{k+1} + w_{n+1}^k \overline{e^{i\varphi}}\right) = \operatorname{Re}\left(w^k, \overline{z^{k+1}}\right).$$

З (3) випливає, що $t_k \operatorname{Re}\left(w^k, \overline{z^{k+1}}\right) \le 0$. Тоді з урахуванням попередніх рівностей і умови (1) отримуємо, що

$$\left\|w^{k+1}\right\|^2 - \left\|w^k\right\|^2 \le \left\|h_\varphi\left(z^{k+1}\right)\right\|^2 \le nc^2 + 1, \quad (k = 0,1,\dots,m)$$

Просумуємо останню нерівність по k від 0 до m. Тоді

$$\left\|w^{m+1}\right\|^2 \le \sum_{k=0}^{m}\left\|h_\varphi\left(z^{k+1}\right)\right\|^2 \le (m+1)\left(nc^2+1\right). \quad (5)$$

Нерівності (4) і (5) суперечливі при великих m. Отже, процес навчання (2) не може тривати нескінченно довго. Теорема доведена.

Як показує лема 2, вибір вагових коефіцієнтів з поля C не приводить до збільшення кількості нейрофункцій. В той же час використання комплексних векторів збільшує витрати пам'яті ЕОМ при побудові нейромереж і кількість операцій при обчисленні зважених сум. Все це приводить до необхідності вивчення НЕ з більш складними знаковими функціями, які б враховували специфіку комплексної площини. Розглянемо один з можливих підходів визначення функції активації. Візьмемо на площині деякий кут. Тоді можна визначити знакову функцію наступним чином:

$$\mathrm{Asign}\, z = \begin{cases} 1, \text{якщо } z \text{ лежить між сторонами кута}, \\ -1, \text{ в іншому випадку}. \end{cases}$$

Не втрачаючи загальності, можна обмежитися розглядом знакових функцій вигляду

$$\mathrm{sign}_\varphi z = \begin{cases} 1, & 0 \le \mathrm{Arg} z \le \varphi, \\ -1, & \text{в іншому випадку}. \end{cases} \quad (6)$$

Бульову функцію $f:\{\alpha,\beta\}^n \to H_2$, для якої для всіх $z \in \{\alpha,\beta\}^n \times \{1\}$ справджується рівність $f(z) = \mathrm{sign}_\varphi(w,\bar{z})$, будемо називати φ-нейрофункцією в алфавіті $\{\alpha,\beta\}$. Клас всіх φ-нейрофункцій в алфавіті $\{\alpha,\beta\}$ позначимо через $P_\varphi(\{\alpha,\beta\},C)$. Покажемо, що клас φ-нейрофункцій є розширенням класу нейрофункцій. Нехай $w_1 = 1-i,\ w_2 = -1+i,\ w_3 = 1+i$. Тоді функція z_1z_2 належить класу $P_{\frac{\pi}{2}}(H_2,C)$, оскільки $z_1z_2 = \mathrm{sign}_{\frac{\pi}{2}}(w,\bar{z})$. Відомо, що функція z_1z_2 не є нейрофункцією. В той же час можна легко показати, що кожна нейрофункція є φ-нейрофункцією.

Лема 3. Для довільних алфавітів $\{\alpha,\beta\}$, $\{\gamma,\delta\}$ має місце взаємно однозначна відповідність $P_\varphi(\{\alpha,\beta\},C) \leftrightarrow P_\varphi(\{\gamma,\delta\},C)$.

Неважко навести приклади, які показують, що лема 2 і теорема 1 не виконуються для φ-нейрофункцій. Наведемо деякі результати, які стосуються φ-нейрофункцій.

Теорема 3. Якщо $0 < \varphi < \pi$, то бульова функція $f(z)$ є φ-нейрофункцією в алфавіті $\{\alpha,\beta\}$ тоді і тільки тоді, коли знайдуться такі нейрофункції $f_1(z)$, $f_2(z)$, що має місце рівність $f(z) = f_1(z) \wedge f_2(z)$.

Теорема 4. Якщо скінченні множини $A_+, A_- \subset C^n \times \{1\}$ можна відокремити за допомогою НЕ зі знаковою функцією виду (5) при деякому $\varphi \in (0,\pi)$, то при довільному $\psi \in (0,\pi)$ їх можна відокремити за допомогою НЕ зі знаковою функцією sign_ψ за допомогою алгоритму навчання, подібного до алгоритму (2).

Список використаних джерел

1. *Айзенберг Н.Н, Иваськив Ю.Л.* Многозначная пороговая логика. – Київ: Наукова думка, 1977.
2. *Уоссермен Ф.* Нейрокомпьютерная техника: теория и практика.– М.:Мир,1992.
3. *Гече Ф.Е., Коцовський В.М.* Оцінка числа бульових функцій, реалізовних на нейронних елементах.// Вісник Ужгородського університету. Серія: фіз.-мат.– 2002. – с. 32-37.
4. *Aizenberg N.N., Aizenberg I.N., Krivosheev G.A.* CNN based on Universal Binary Neurons: Learning algorithm with Error-Correction and Application to Impulsive-Noise Filtering on Gray-Scale Images // Proceedings of the Fourth International Workshop on Cellular Neural Networks and their Applications, Seville, Spain, June 24-26, 1996. – P. 309-314.
5. *Айзенберг И. Н.* Универсальный пороговый элемент над полем комплексных чисел // Кибернетика. – 1991. – № 3 с. 272-285.
6. *Заенцев И. В.* Нейронные сети: основные модели. – Воронеж: изд. ВГУ, 1999.
7. *Горбань А.Н., Россиев Д.А.* Нейронные сети на персональном компьютере. – Новосибирск: Наука, 1996.

Надійшла до редакції 10.10.2006