УДК 681.14

Ф.Е. Гече, В.М. Коцовський (Ужгородський нац. ун-т), А.Є. Батюк (Львівський політехнічний ін-т)

АЛГОРИТМИ НАВЧАННЯ УЗАГАЛЬНЕНИХ НЕЙРОННИХ ЕЛЕМЕНТІВ ВІДНОСНО СИСТЕМИ ХАРАКТЕРІВ.

В роботі розглядаються узагальненні нейронні елементи відносно системи характерів, наводяться алгоритми навчання цих елементів та доводиться їх збіжність.

Generalized neural devices over the character set have been studied in the paper. Learning algorithms have been given and their convergence has been proved in the paper.

Нейронні елементи (НЕ) використовуються при побудові нейромереж, які є ефективним механізмом для розв'язування широкого кола задач класифікації об'єктів, розпізнавання образів, стиску інформації, прогнозування поведінки динамічних систем, наближення і екстраполяції функцій [1]. При цьому застосовуються різноманітні підходи по визначенню конфігурації НЕ і зв'язків між ними у нейромережах. У роботі вивчаються НЕ відносно системи характерів, розгляд яких дозволяє значно розширити клас бульових нейрофункцій.

Нехай $H_2 = \{-1,1\}$, $G_n = H_2 \times \ldots \times H_2$ – n-а декартова степінь множини H_2. Дискретні функції вигляду $f : G_n \to H_2$ будемо називати бульовими функціями в алфавіті $\{-1,1\}$. Визначимо відображення $\chi_j : G_n \to H_2$ наступним чином $\chi_j(a) = a_1^{j_1} \cdot a_2^{j_2} \cdot \ldots \cdot a_n^{j_n}$, де $a = (a_1, \ldots, a_n) \in G_n$, $j = j_1 2^{n-1} + j_2 2^{n-2} + \ldots + j_n$ $\left(j_i \in \{0,1\},\ i = \overline{1,n}\right)$. Тобто $\chi_j(a) = a_{l_1} \cdot \ldots \cdot a_{l_r}$, при умові, що $j = 2^{n-l_1} + 2^{n-l_2} + \ldots + 2^{n-l_r}$, $1 \le l_1 < l_2 < \ldots < l_r \le n$, $r \le n$. У відповідності до [2] відображення χ_j, $j = 0,1,\ldots,2^n - 1$ будемо називати характерами групи G_n над полем R. На множині $R \setminus \{0\}$ визначимо знакову функцію Rsign так:

$$\text{Rsign } x = \begin{cases} 1, & \text{якщо } x > 0, \\ -1, & \text{якщо } x < 0. \end{cases}$$

Нехай $X = \{\chi_{i_1}, \ldots, \chi_{i_m}\}$ – деяка система характерів, $w = (w_1, \ldots, w_m) \in R^m$. Якщо для бульової функції $f : G_n \to H_2$ виконується умова

$$\text{для всіх } a = (a_1, \ldots, a_n) \in G_n \quad f(a) = \text{Rsign} \sum_{j=1}^{n} w_j \chi_{i_j}(a), \tag{1}$$

то будемо казати, що відносно системи характерів X бульова функція f реалізується на НЕ з ваговим вектором w. Якщо для бульової функції f можна вказати принаймні один такий ваговий вектор w, для якого виконується умова (1), то функцію f будемо називати X-нейрофункцією. При цьому для вагового вектора w має виконуватися умова для всіх $a \in G_n$ $(w, \chi(a)) \ne 0$, де $(w, \chi(a)) = w_1 \chi_{i_1}(a) + \ldots + w_m \chi_{i_m}(a)$ – скалярний добуток векторів w і $\chi(a) = (\chi_1(a), \ldots, \chi_m(a))$. Такі вагові вектори будемо називати X-допустимими. Зауважимо, що поняття X-нейрофункції є узагальненням поняття порогової бульової функції. Справді, якщо покласти $X = \{\chi_0, \chi_1, \chi_2, \chi_4, \ldots, \chi_{2^{n-1}}\}$, то отримуємо клас "звичайних" порогових функцій. Позначимо через $W_X(f)$ множину вагових векторів всіх НЕ, на яких реалізується X-нейрофункція f.

Під процесом навчання НЕ для бульової X-нейрофункції f будемо розуміти процес побудови скінченої послідовності векторів

$$w^0, w^1, \ldots, w^t, \tag{2}$$

такої, що функція *f* реалізується на НЕ з ваговим вектором w^t. В роботі розглядаються та обґрунтовуються два методи побудови послідовності (2).

1. Спектральний алгоритм навчання НЕ.

Для бульової функції $f : G_n \to H_2$ відносно системи характерів *X* визначимо характеристичний вектор $s_X(f) = (s_1, \dots s_m)$ так: $s_j = \sum_{a \in G_n} \chi_{i_j}(a) f(a), \quad j = \overline{1, m}$. Компоненти вектора s_f відрізняються від відповідних спектральних коефіцієнтів функції *f* [3] лише на сталий множник 2^n. Будемо припускати, що функція *f* є *X*-нейрофункцією. Нехай $w \in W_X(f) \subset R^m$ – довільний ваговий вектор НЕ, який реалізує функцію *f*. Будемо вважати, що перші *k*+1 членів $w^0, w^1, \dots, w^k$ послідовності (2) вже відомі і нехай w^k – *m*-вимірний *X*-допустимий дійсний вектор. Якщо функція *f* реалізується на НЕ з ваговим вектором w^k, то процес навчання вважаємо завершеним. Припустимо, що *f* не реалізується на НЕ з ваговим вектором w^k. Опишемо алгоритм, за яким можна побудувати такий вектор w^{k+1}, для якого справджується нерівність

$$\left\| w^{k+1} - w \right\| < \left\| w^k - w \right\|, \tag{3}$$

(де $\|x\| = \sqrt{(x, x)}$ – звичайна евклідова норма в просторі R^m). Нехай $z^k = w^{k+1} - w^k$ – вектор приросту. Оскільки умова (3) рівносильна умові $\left\| w^{k+1} - w \right\|^2 - \left\| w^k - w \right\|^2 < 0$, то для виконання умови (3) досить вимагати виконання умови

$$2\left(z^k, w - w^k\right) - \left\| z^k \right\|^2 > 0. \tag{4}$$

Переходимо до визначення вектора z^k. Нехай $f^k(x_1, \dots x_n)$ – функція, яка відносно системи характерів *X* реалізується на НЕ з ваговим вектором w^k. Із зроблених припущень випливає, що $f^k \neq f$ і $s_X\left(f^k\right) \neq s_X(f)$. Вектор приросту z^k будемо шукати у вигляді

$$z^k = \alpha_k \left(s_X(f) - s_X\left(f^k\right) \right), \tag{5}$$

де α_k – деяка додатна скалярна величина. Підставимо у (4) значення z^k з (5). Отримаємо

$$2\alpha_k \left(w - w^k, s_X(f) - s_X\left(f^k\right) \right) - \alpha_k^2 \left\| s_X(f) - s_X\left(f^k\right) \right\|^2 > 0. \tag{6}$$

Відомо [4], що якщо $f^k \neq f$, то

$$\left(w, s_X(f)\right) - \left(w, s_X\left(f^k\right)\right) > 0 \text{ і } \left(w^k, s_X(f)\right) - \left(w^k, s_X\left(f^k\right)\right) < 0. \tag{7}$$

Тому перший доданок у рівності (6) є строго додатним. Ліва частина (6) є квадратним тричленом відносно α_k, який приймає найбільше значення при $\alpha_k = \dfrac{\left(w - w^k, s_X(f) - s_X\left(f^k\right) \right)}{\left\| s_X(f) - s_X\left(f^k\right) \right\|^2}$ і це значення є додатним. Але останнє співвідношення не може бути застосоване для знаходження α_k, оскільки воно містить невідому величину *w*. Покажемо, що нерівність (6) виконується, якщо в якості α_k вибрати довільне число вигляду $t_k \alpha_k^0$, де

$$\alpha_k^0 = \frac{\left(w^k, s_X\left(f^k\right) - s_X(f) \right)}{\left\| s_X(f) - s_X\left(f^k\right) \right\|^2},$$

а величина t_k задовольняє умову

$$0 < t_k < 2\left(1 + \frac{\left(w, s_X(f) - s_X\left(f^k\right) \right)}{\left(w^k, s_X\left(f^k\right) - s_X(f) \right)} \right). \tag{8}$$

Число t_k надалі будемо називати нормуючим множником приросту.

З (7) випливає, що $t_k\alpha_k^0>0$. Після підстановки у нерівність (6) замість α_k величини $t_k\alpha_k^0$ і елементарних перетворень, отримуємо наступну нерівність

$$\frac{\left(2t_k-t_k^2\right)\left(w^k,s_X\left(f^k\right)-s_X(f)\right)^2+2t_k\left(w,s_X(f)-s_X\left(f^k\right)\right)\cdot\left(w^k,s_X\left(f^k\right)-s_X(f)\right)}{\left\|s_X(f)-s_X\left(f^k\right)\right\|^2}>0,$$

яку можна розглядати як квадратну нерівність відносно невідомої t_k. Остання нерівність виконується при умові, що $0<t_k<2\left(1+\left(w,s_X(f)-s_X\left(f^k\right)\right)\cdot\left(w^k,s_X\left(f^k\right)-s_X(f)\right)^{-1}\right)$. Тому нерівність (6) також виконується. Твердження доведене.

Надалі обмежимося розглядом векторів приросту вигляду

$$z^k=t_k\alpha_k^o\left(s_X(f)-s_X\left(f^k\right)\right), \tag{9}$$

де нормуючий множник t_k задовольняє умову (8).

Зауваження 1. Для приростів виду (9) нерівність (3) виконується для всіх вагових векторів $w\in W_X(f)$ НЕ, на яких реалізується функція *f*, при умові, що нормуючий множник приросту t_k задовольняє умову $0<t_k\le 2$.

Умова (8) задає межі для допустимих значень нормуючого множника приросту t_k. Слід зазначити, що у (8) величина t_k обмежена зверху сумою $2+\dfrac{2\left(w,s_X(f)-s_X\left(f^k\right)\right)}{\left(w^k,s_X\left(f^k\right)-s_X(f)\right)}$, в якій у другий доданок входить невідомий наперед скалярний добуток $\left(w,s_X(f)-s_X\left(f^k\right)\right)$, про який у загальному випадку відомо лише те, що згідно (7) він набуває строго додатне значення.

Зауваження 2. При виборі вектора приросту z^k у вигляді (9) необхідно вимагати, щоб для всіх $a\in G_n$ $\left(w^{k+1},\chi(a)\right)\neq 0$, бо інакше вектор w^{k+1} не є *X*-допустимим і на НЕ з таким ваговим вектором не можна реалізувати жодну цілком визначену бульову функцію. Цього можна завжди досягти за рахунок незначних змін вектора приросту, змінивши відповідним чином нормуючий множник приросту t_k так, щоб він не вийшов за межі допустимих значень (8).

Розглянемо деякі властивості, які має послідовність вагових векторів $\left\{w^k\right\}$, побудована по правилу (5).

Покажемо, що якщо $t_k\ge 1$, то виконується нерівність $s_X\left(f^{k+1}\right)\neq s_X\left(f^k\right)$. Припустимо протилежне. Тоді з урахуванням (7) маємо $\left(w^{k+1},s_X\left(f^k\right)-s_X(f)\right)>0$. Замінивши w^{k+1} на $w^k+t_k\dfrac{\left(w^k,s_X\left(f^k\right)-s_X(f)\right)}{\left\|s_X\left(f^k\right)-s_X(f)\right\|^2}\cdot\left(s_X(f)-s_X\left(f^k\right)\right)$, отримаємо $\left(w^k,s_X\left(f^k\right)-s_X(f)\right)\left(1-t_k\right)>0$. З останньої нерівності випливає, що $t_k<1$. Тоді отримуємо протиріччя з умовою $t_k\ge 1$.

Зауваження 3. Доведена властивість не виключає того, що знайдуться такі натуральні числа *k* і *l* $(k<l)$, для яких має місце рівність $s_X\left(f^k\right)=s_X\left(f^l\right)$. В такому випадку по аналогії до [5] будемо казати, що в області характеристичних векторів утворився граничний цикл. Будемо вважати, що цей цикл не має власних граничних підциклів. Нехай

$$S\cdot y=0$$

система *m* рівнянь з $l-k$ невідомими $y_1,\ldots,y_{l-k}$ при умові, що стовпцями матриці *S* є вектори $s_X(f)-s_X\left(f^{k+i-1}\right)$, $i=\overline{1,l-k}$. Якщо попередня система має додатний розв'язок

$\left(y_i \geq 0,\ i=\overline{1,l-k},\ \sum_{i=1}^{l-r} y_i > 0\right)$, то функція f не є X-нейрофункцією. Дійсно, якщо припустити, що функція f реалізується на НЕ з ваговим вектором w, то домноживши обидві частини рівності $\sum_{i=1}^{l-k} y_i\left(s_X(f)-s_X\left(f^{k+i-1}\right)\right)=0$ скалярно на w, отримаємо $\sum_{i=1}^{l-k} y_i\left(s_X(f)-s_X\left(f^{k+i-1}\right), w\right)=0$. Остання рівність суперечить нерівностям (7). Якщо система не має додатних розв'язків, то процес навчання продовжуємо.

Прослідкуємо тепер, як змінюється $\left\|w^{k+1}\right\|$ порівняно з $\left\|w^k\right\|$, якщо вектор приросту z^k шукається згідно правила (9) і нормуючий множник t_k задовольняє (8). Легко бачити, що для норм має місце рівність $\left\|w^{k+1}\right\|^2=\left\|w^k\right\|^2-2t_k\frac{\left(w^k, s_X\left(f^k\right)-s_X(f)\right)^2}{\left\|s_X\left(f^k\right)-s_X(f)\right\|^2}+t_k^2\frac{\left(w^k, s_X\left(f^k\right)-s_X(f)\right)^2}{\left\|s_X\left(f^k\right)-s_X(f)\right\|^2}$, з якої випливає, що $\left\|w^{k+1}\right\|<\left\|w^k\right\|$, якщо $t_k<2$, $\left\|w^{k+1}\right\|=\left\|w^k\right\|$, якщо $t_k=2$ і $\left\|w^{k+1}\right\|>\left\|w^k\right\|$, якщо $t_k \geq 2$.

Для обґрунтування спектрального алгоритму навчання нам знадобиться

Теорема 1. *Нехай f – бульова функція, w^0 – X-допустимий ваговий вектор, прирости для послідовністі $\left\{w^k\right\}$ визначаються згідно* (5)*, нормуючий множник t_k вибирається з проміжку $\left[t_0, 2\right]$, де $0<t_0<2$, $f^k \neq f$, $k=0,1,\ldots$. Тоді якщо знайдеться таке $\varepsilon > 0$, що для всіх $k \in N$ і для всіх $a \in G_n$ виконується умова*

$$\left|\left(w^k, \chi(a)\right)\right| \geq \varepsilon, \tag{10}$$

то функція f не є X-нейрофункцією.

Доведення. Припустимо противне. Тоді f – X-нейрофункція, для якої процес навчання триває як завгодно довго і для послідовності $\left\{w^k\right\}$ виконується умова (10). З вибору нормуючих множників t_k і зауваження 1 випливає, що для всіх $w \in W_X(f)$ виконуються нерівності $\left\|w^{k+1}-w\right\|<\left\|w^k-w\right\|$, $k=0,1,\ldots$. Тоді на основі [4] робимо висновок, що послідовність $\left\{w^k\right\}$ збігається до деякого вектора w^*. Покажемо, що $w^* \in W_X(f)$. Із нерівності (10) і неперервності скалярного добутку випливає, що знайдеться такий номер l, що для всіх $k \geq l$ і для всіх $a \in G_n$ виконується умова $\mathrm{Rsign}\left(w^k, a\right)=\mathrm{Rsign}\left(w^*, a\right)$. Тоді вектор w^* є X-допустимим і НЕ з ваговим вектором w^* реалізує бульову функцію f^*, причому виконуються рівності $f^*=f^l=f^{l+1}=\ldots$ і $s_X\left(f^*\right)=s_X\left(f^l\right)=s_X\left(f^{l+1}\right)=\ldots$. Якщо всі $t_k \geq 1$, то отримуємо протиріччя з нерівністю $s_X\left(f^{k+1}\right) \neq s_X\left(f^k\right)$. Доведемо, що у випадку $0<t_k<1$ $f=f^*=f^l$. Зі збіжності послідовності $\left\{w^k\right\}$ випливає, що вектор приросту $z^k=t_k\frac{\left(w^k, s_X\left(f^k\right)-s_X(f)\right)}{\left\|s_X\left(f^k\right)-s_X(f)\right\|^2}\left(s_X(f)-s_X\left(f^k\right)\right) \to 0$. Тому $\left(w^k, s_X\left(f^k\right)-s_X(f)\right) \to 0$. Тому і $\left(w^*, s_X\left(f^*\right)-s_X(f)\right)=0$ Тоді з нерівності (7) випливає, що $s_X(f)=s_X\left(f^*\right)$. Тому $f^l \in W_X(f)$. Отримуємо протиріччя з припущенням, що $f^k \neq f$, $k=0,1,\ldots$ Теорема доведена.

Зауваження 4. Якщо бульова функція f є X-нейрофункцєю і для послідовності $\left\{w^k\right\}$ умова (10) не виконується, то не можна стверджувати, що $w^* \in W_X(f)$. В цьому випадку може виявитися, що вектор w^* уже не є X-допустимим і належить замиканню множини $W_X(f)$.

На основі теореми 1 можна запропонувати таку модифікацію алгоритму навчання при умові, що нормуючий множник t_k вибирається з проміжку $\left[t_0,2\right]$. Вибираємо деяке достатньо мале число $\varepsilon > 0$ і здійснюємо процес навчання до тих пір, поки не почне виконуватися одна з наступних умов:

1. $s_X\left(f^k\right)=s_X\left(f\right)$. Тоді робимо висновок, що бульова функція f відносно системи характерів X реалізується на НЕ з ваговим вектором w^k і процес навчання завершено.
2. В області характеристичних векторів утворився граничний цикл. Тоді діємо згідно зауваження 3.
3. Порушилася нерівність (10). Тоді припиняємо процес навчання або вносимо в нього якісь зміни (змінюємо початкове наближення, нормуючі множники t_k на кількох останніх кроках алгоритму або зменшуємо ε).

1. Уоссермен Ф. Нейрокомпьютерная техника: теория и практика.– М.:Мир,1992.
2. Ван дер Варден Б. Л. Алгебра. – М.: Наука, 1979.
3. Айзенберг Н.Н., Иваськив Ю.Л. Многозначная пороговая логика. – Київ: Наукова думка, 1977.
4. Грицик В.В., Гече Ф.Е. Реалізація бульових та багатозначних бульових функцій на нейронних елементах // Доповіді НАН України №5, 2004, С. 65-68.
5. Дертоузос М. Пороговая логика. – М.: Мир, 1967.