УДК 519.816(045)

О.Є. ЛИТВИНЕНКО
Київський національний авіаційний університет, Київ, Україна,
e-mail: *litvinen@nau.edu.ua.*

Д.П. КУЧЕРОВ
Київський національний авіаційний університет, Київ, Україна,
e-mail: *d_kucherov@ukr.net.*

М.М. ГЛИБОВЕЦЬ
Національний університет «Києво-Могилянська академія», Київ, Україна,
e-mail: *glib@ukma.edu.ua.*

ДЕКОМПОЗИЦІЙНИЙ МЕТОД ОБЧИСЛЕННЯ ВАГОВИХ КОЕФІЦІЄНТІВ БІНАРНОЇ НЕЙРОННОЇ МЕРЕЖІ

Анотація. Викладено метод обчислення вагових коефіцієнтів бінарної нейронної мережі на основі її декомпозиції на елементарні модулі. Метод дає змогу обчислювати вагові коефіцієнти всіх зв'язків мережі на етапі її проєктування, в результаті чого немає потреби у реалізації трудомістких ітераційних алгоритмів навчання мережі в процесі її експлуатації. Наведено алгоритм та приклад обчислення вагових коефіцієнтів.

Ключові слова: бінарна нейронна мережа, вагові коефіцієнти, метод обчислення, декомпозиція, алгоритм.

ВСТУП

Широке застосування нейронних мереж (НМ) для розв'язання різних складних прикладних задач зумовлено сучасними досягненнями у сфері штучного інтелекту, обчислювальної техніки та програмування. Основною перевагою використання моделі НМ є природність реалізації їхньої здатності до самонавчання та розпаралелювання [1, 2].

Через велику розмірність моделей НМ існує супутня низка їхніх недоліків. Серед них традиційно виокремлюють: використання евристики в більшості підходів до проєктування, тривалість ітераційного налаштування внутрішніх зв'язків між базисними елементами, проблеми із пошуком реальних даних для формування навчальної вибірки, можливість отримання хибного розв'язку тощо. Частково зазначені проблеми розв'язують покращенням структури НМ, поділом за часом процесів навчання та експлуатації моделі [3–5].

У цій статті запропоновано один із підходів до реалізації бінарної нейронної мережі, входами та виходами якої є набори бітів, а нейрони реалізують функції двійкової логіки кількох змінних. Мережі цього типу застосовують для розв'язання задач кластеризації у системах комп'ютерного зору, телекомунікаційних пристроях оброблення сигналів, системах діагностування захворювань тощо. У реальних задачах розмірність входу може досягати великих розмірів.

Запропонований підхід суттєво відрізняється від побудови мереж відомого типу. Очікуються кращі показники НМ за рахунок кінцевої кількості варіантів повного перебору функцій, отже, навчання закінчуватиметься за прийнятний час і апаратна реалізація мережі буде набагато простішою.

ОГЛЯД ЛІТЕРАТУРИ

Однією із головних задач проєктування штучної НМ є обчислення вагових коефіцієнтів її зв'язків [1–5]. Для її розв'язання зазвичай використовують підхід, що передбачає реалізацію процедури «навчання» НМ [6, 7].

Алгоритми навчання НМ базуються на двох правилах Д. Хебба [3–6], які для бінарних мереж можна сформулювати так:

— якщо вихідний сигнал нейрона неправильний і дорівнює 0, потрібно збільшити вагові коефіцієнти його вхідних зв'язків, якими пройшов сигнал, рівний 1;

— якщо вихідний сигнал нейрона неправильний і дорівнює 1, потрібно зменшити вагові коефіцієнти його вхідних зв'язків, якими пройшов сигнал, рівний 1.

Найбільш поширеним методом навчання НМ «з учителем» є метод зворотного розповсюдження помилки згідно з «дельта-правилом» [1–6].

Для навчання НМ «без учителя» найчастіше застосовують метод корекції помилки згідно з «альфа-правилом» [1–7]. Цей метод передбачає покрокове збільшення вагових коефіцієнтів активних зв'язків НМ на однакову (заздалегідь обчислену) величину зі збереженням колишніх значень вагових коефіцієнтів зв'язків, які не є активними.

Синтез та апаратна реалізація нейроподібних мереж досліджені у [8, 9].

Можливість побудови НМ на дискретних елементах з елементами пам'яті, бінарним введенням та виведенням описана в [10] для моделювання життя популяції віртуальних бактерій, що розвиваються в обмеженій двовимірній області.

У [11] розглянуто навчання одношарової та багатошарової нейронної мережі з коротким зв'язком з використанням бінарної класифікації за умови досягнення нульової помилки навчання у всіх локальних мінімумах із правильно обраною сурогатною функцією втрат.

Задача бінарної класифікації не є тривіальною у випадку, якщо об'єктами є супутникові [10] або рентгенівські знімки [13, 14].

У [15] наведено схему пошуку зображень, засновану на атрибутах із CNN, що використовує ефективний двійковий автокодувальник невеликого розміру та процедуру пошуку ближнього сусіда.

У [16] пропонують розв'язувати проблему залежності між точністю та продуктивністю навчання бінарного класифікатора баєсівського типу за допомогою компромісу. Для цього вводиться функція ризику, в якій вплив однієї з досліджуваних характеристик компенсується запровадженням відповідного множника Лагранжа.

Зазначимо, що всі відомі алгоритми навчання НМ мають ітераційний характер і вимагають значних витрат машинного часу. Однак для багатьох прикладних областей існує можливість апріорного обчислення вагових коефіцієнтів НМ на етапі проєктування, що дає змогу уникнути трудомісткої ітераційної процедури навчання в процесі експлуатації.

Метою цієї статті є викладення підходу до розв'язування задачі апріорного обчислення вагових коефіцієнтів бінарної НМ.

ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

Припустимо, що на вхід класифікаційної НМ надходить вектор вхідних сигналів $x=(x_i \mid i=\overline{1,n_1})$, $x\in X$, а на виході генерується вектор вихідних сигналів $y=(y_j \mid j=\overline{n-n_r+1,n})$, $y\in Y$, де X — множина можливих вхідних векторів; Y — множина вихідних векторів; n — кількість нейронів у НМ; i, j, — номера нейронів, $i=\overline{1,n}$, $j=\overline{1,n}$; r — кількість шарів НМ, яка визначається вимогами до точності розв'язку задачі.

Передбачається також, що НМ є бінарною, тобто вхідні і вихідні сигнали всіх нейронів можуть набувати значень з множини $\{0,1\}$:

$$(\forall i=\overline{1,n_1})(x_i\in\{0,1\});\ (\forall k=\overline{2,r})(i\in I_{k-1})(x_i\in\{0,1\});\ (\forall j=\overline{1,n})(y_j\in\{0,1\}),$$

де I_k — множина номерів нейронів, що належать k-му шару, $k=\overline{1,r}$; k — номер шару нейронної мережі, $k=\overline{1,r}$; n_k — кількість нейронів, що належать k-му шару, $k=\overline{1,r}$.

Функції активації $f_j(s_j)$ всіх нейронів є одиничними ступінчастими функціями з нульовими усуненнями.

Для нейронів першого (вхідного) шару $f_i(s_i) = x_i$, $i = \overline{1, n_1}$. Для нейронів всіх наступних шарів

$$f_j(s_j) = \begin{cases} 0 \text{ для } s_j \le 0, \\ 1 \text{ для } s_j > 0, \end{cases}$$

де s_j — виважена сума вхідних сигналів j-го нейрона, $s_j = \sum_{i \in I_{k-1}} w_{ij} x_{ij}$, $j \in I_k$, $k = \overline{2, r}$. Тут x_{ij} — вхідний сигнал j-го нейрона, що надходить від i-го нейрона попереднього шару НМ, $x_{ij} = y_i$, $i \in I_k$, $j \in I_{k+1}$, $k = \overline{1, r-1}$; w_{ij} — ваговий коефіцієнт зв'язку, що йде від i-го нейрона до j-го нейрона, $i \in I_k$, $j \in I_{k+1}$, $k = \overline{1, r-1}$.

Якщо кожному можливому вхідному вектору з навчальної вибірки НМ ставиться у відповідність очікуваний (еталонний) вихідний вектор, то НМ правильно класифікує вхідний сигнал. Тож критерій якості бінарної НМ можна визначити виразом

$$\varepsilon = \frac{1}{L} \sum_{l=1}^{L} (y_l - y^*)^2 \le \varepsilon_{\text{доп}},$$

де L — кількість циклів налаштування НМ; l — номер циклу налаштування; ε — середньоквадратична похибка класифікації НМ за прийнятою навчальною вибіркою векторів $x \in X$; y_l — вектор значень вихідних сигналів у l-му циклі налаштування; y^* — вектор еталонних значень вихідних сигналів, $y^* \in Y$; $\varepsilon_{\text{доп}}$ — деяке додатне дійсне число, що визначається конструктором НМ і характеризує допустиме відхилення фактичного вихідного сигналу від еталонного.

Потрібно обчислити вектор значень вагових коефіцієнтів $w = (w_{ij} \mid i \in I_k$, $j \in I_{k+1}$, $k = \overline{1, r-1})$ таких, що під час надходження на вхід НМ вектора бінарних сигналів $x \in X$ мережа буде видавати вектор правильних значень вихідних бінарних сигналів $y \in Y$, що задовольняють критерій ε.

ОБЧИСЛЕННЯ ВАГОВИХ КОЕФІЦІЄНТІВ

Особливістю НМ є взаємна незалежність вагових коефіцієнтів зв'язків між нейронами. Інакше кажучи, вагові коефіцієнти $w_{ij'}$ $(i \in I_k, j' \in I_{k+1})$ зв'язків, що належать j'-му нейрону, не пов'язані жодними математичними співвідношеннями з ваговими коефіцієнтами $w_{ij''}$ $(i \in I_k, j'' \in I_{k+1}, j' \ne j'')$ зв'язків, що належать j''-му нейрону, $k = \overline{1, r-1}$. Це надає можливість провести декомпозицію НМ на прості фрагменти (елементарні модулі) і розв'язувати задачу обчислення вагових коефіцієнтів для кожного модуля окремо.

Пояснимо суть запропонованого методу.

Для кожного набору значень вхідних сигналів $x = (x_i \mid i = \overline{1, n_1})$ встановлюють відповідний набір еталонних (правильних) значень вихідних сигналів НМ $y^* = (y_j^* \mid j = \overline{n - n_r + 1, n})$.

Далі на мережі виокремлюють фрагменти (елементарні модулі), що складаються з одного нейрона k-го шару N_j, $j \in I_k$, та всіх нейронів попереднього $(k-1)$-го шару N_i, $i \in I_{k-1}$, $k = \overline{1, r-1}$ (рис. 1).

Виокремлення елементарних модулів здійснюють послідовно, починаючи з вихідного шару НМ $(k = r)$ і закінчуючи другим шаром $(k = 2)$.

Далі окремо для кожного j-го $(j = \overline{n - n_r + 1, n})$ вихідного каналу мережі обчислюють значення вагових коефіцієнтів w_{ij}, $i \in I_{r-1}$, зв'язків, що йдуть до j-го нейрона від нейронів попереднього $(r-1)$-го шару.

Вагові коефіцієнти w_{ij}, $i \in I_{r-1}$, повинні бути такими, щоб вихідний сигнал j-го нейрона збігався з потрібним еталонним значенням $y_j = y_j^*$.

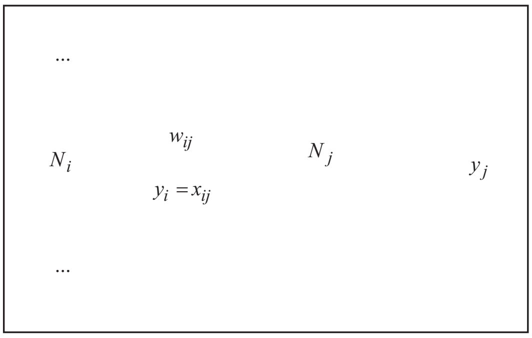

Рис. 1. Елементарний модуль НМ

Вихідний сигнал y_j має значення функції активації $f_j(s_j)$. Отже, функція активації повинна набувати значення y_j^*: $f_j(s_j) = y_j^*$.

Якщо $y_j^* = 0$, то j-й нейрон не повинен бути активованим, отже, зважена сума його вхідних сигналів не повинна перевищувати 0:

$$s_j = \sum_{i \in I_{r-1}} w_{ij} x_{ij} \le 0. \tag{1}$$

Якщо $y_j^* = 1$, то j-й нейрон повинен бути активованим, отже, зважена сума його вхідних сигналів повинна бути більше 0:

$$s_j = \sum_{i \in I_{r-1}} w_{ij} x_{ij} > 0. \tag{2}$$

Вирази (1) і (2) дають змогу сформулювати умови, яким повинні задовольняти значення вагових коефіцієнтів w_{ij}, $i \in I_{r-1}$. Для цього потрібно послідовно підставити у формули (1) та (2) всі можливі комбінації значень вхідних сигналів $x_{ij} \in \{0,1\}$, $i \in I_{r-1}$.

Нехай $x_j^* = (x_{ij}^* \mid i \in I_{r-1})$ — вектор значень вхідних сигналів, що надходять на вхід j-го нейрона, $j \in I_r$; $I_{r-1}^{(1)}$ — множина номерів нейронів $(r-1)$-го шару, від яких на вхід j-го нейрона надходять сигнали, рівні 1, $I_{r-1}^{(1)} = \{i \in I_{r-1} \colon x_{ij}^* = 1\}$. Тоді умови (1), (2), яким повинні задовольняти значення вагових коефіцієнтів w_{ij}, $i \in I_{r-1}$, можна записати у вигляді таких нерівностей:

$$\sum_{i \in I_{r-1}^{(1)}} w_{ij} \le 0 \quad \text{для} \quad y_j^* = 0 \quad \text{та} \quad \sum_{i \in I_{r-1}^{(1)}} w_{ij} > 0 \quad \text{для} \quad y_j^* = 1.$$

Для бінарних мереж область значень вагових коефіцієнтів достатньо обмежити множиною $\{0, 1\}$. При цьому умови (1), (2) можна записати такими імплікативними виразами:

$$(y_j^* = 0) \to \left(\sum_{i \in I_{r-1}^{(1)}} w_{ij} = 0 \right); \; (y_j^* = 1) \to \left(\sum_{i \in I_{r-1}^{(1)}} w_{ij} \ge 1 \right).$$

Нерівності, що входить до другого виразу, задовольняє множина комбінацій значень вагових коефіцієнтів w_{ij}, $i \in I_{r-1}^{(1)}$. Щоб уникнути невизначеності, доцільно прийняти правило, згідно з яким за умови $y_j^* = 1$ всім ваговим коефіцієнтам привласнюють значення 1. У підсумку правила обчислення вагових коефіцієнтів w_{ij}, $i \in I_{r-1}^{(1)}$, бінарної мережі можна записати так:

$$(y_j^* = 0) \to [(\forall i \in I_{r-1}^{(1)})(w_{ij} = 0)], \tag{3}$$

$$(y_j^* = 1) \to [(\forall i \in I_{r-1}^{(1)})(w_{ij} = 1)]. \tag{4}$$

Загальна формула обчислення вагових коефіцієнтів бінарної НМ має вигляд

$$w_{ij} = \begin{cases} x_{ij}, & \text{якщо } y_j^* = x_{ij}, \\ (1 - x_{ij}), & \text{якщо } y_j^* \ne x_{ij}, \end{cases} \tag{5}$$

$$y_j^* \in \{0,1\}, \; x_{ij} \in \{0,1\}, \; j = \overline{1,n}, \; i \in I_j,$$

де I_j — множина номерів нейронів, від яких надходять сигнали на вхід j-го нейрона (інакше кажучи, множина номерів нейронів шару, що є попереднім для того шару, в якому міститься j-й нейрон).

З формули (5) випливають два висновки, які визначають необхідні та достатні умови отримання еталонних значень вихідного сигналу елементарного модуля бінарної мережі:

— для отримання значення $y_j^* = 0$ потрібно, щоб вагові коефіцієнти всіх вхідних зв'язків j-го нейрона дорівнювали нулю $(\forall i \in I_j)(w_{ij} = 0) \to (y_j^* = 0)$;

— для отримання значення $y_j^* = 1$ потрібно, щоб серед вхідних сигналів j-го нейрона був хоча би один сигнал $x_{ij} = 1$, що надходить по вхідному зв'язку, ваговий коефіцієнт якого більше нуля $(\exists i \in I_j)[(x_{ij} = 1) \& (w_{ij} > 0)] \to (y_j^* = 1)$.

Після обчислення вагових коефіцієнтів вхідних зв'язків j'-го нейрона r-го шару описана процедура реалізується для наступного j-го $(j \in J_r \setminus \{j'\})$ нейрона вихідного шару мережі.

Після обчислення вагових коефіцієнтів вхідних зв'язків всіх нейронів r-го шару формуються елементарні модулі, кожен з яких складається з одного нейрона $(r-1)$-го шару та всіх нейронів попереднього $(r-2)$-го шару. Далі окремо для кожного модуля наведеним способом обчислюють значення вагових коефіцієнтів зв'язків між нейронами $(r-2)$-го та $(r-1)$-го шарів, після чого розглядають попередню пару шарів: $(r-3)$-го та $(r-2)$-го і т. д. При цьому вихідний сигнал y_j $(j \in I_k)$ кожного модуля почергово приймається рівним 0 та 1, а як вхідні сигнали x_{ij}, $i \in I_{k-1}$, розглядаються всі можливі бінарні комбінації їхніх значень, $2 \le k \le r-1$.

Процес завершується обчисленням вагових коефіцієнтів вхідних зв'язків між нейронами першого і другого шарів НМ. У результаті визначається повний вектор $w = (w_{ij} \mid i \in I_k,\ j \in I_{k+1},\ k = \overline{1, r-1})$ вагових коефіцієнтів, який забезпечує отримання еталонних значень вихідних сигналів для будь-яких комбінацій сигналів, що можуть надходити на вхід НМ.

Під час проєктування мережі слід передбачати додатковий вхідний сигнал x_0, призначений для ініціалізації мережі у випадку надходження на її вхід нульового вектора $x = (x_i = 0 \mid i = \overline{1, n_1})$: $x_0 = \prod_{i=1}^{n_1} (1 - x_i)$.

АЛГОРИТМ ОБЧИСЛЕННЯ ВАГОВИХ КОЕФІЦІЄНТІВ

Алгоритм обчислення вагових коефіцієнтів міжнейронних зв'язків НМ передбачає послідовне виконання таких дій.

1. Фіксація номера j нейрона N_j, який на поточній ітерації буде розглядатися як вихідний нейрон елементарного модуля. Значення j вибирається з множини $\{n - n_1 - 1, \ldots, n\}$ у порядку спадання, починаючи з n і закінчуючи $(n - n_1 - 1)$.

2. Формування множини I_j номерів нейронів, які безпосередньо передують нейрону N_j.

3. Вибір еталонного значення вихідного сигналу j-го нейрона y_j^* з множини $\{0, 1\}$.

4. Формування множин $M_j^{(z)}$, $z \in \{0, 1\}$, векторів $x_i = (x_0,\ x_{ij} \mid i \in I_j)$ значень вхідних сигналів нейрона N_j, за яких він повинен видавати еталонний вихідний сигнал y_j^*.

5. Обчислення вагових коефіцієнтів w_{ij}, $i \in I_j$, вхідних зв'язків нейрона N_j за формулою (5).

6. Перевірка умов завершення циклів алгоритму.

Якщо вагові коефіцієнти w_{ij}, $i \in I_j$, обчислені не для всіх комбінацій вхідних сигналів $x_i = (x_0,\ x_{ij} \mid i \in I_j)$, то перехід до п. 4.

Якщо вагові коефіцієнти w_{ij}, $i \in I_j$, обчислені для всіх комбінацій вхідних сигналів, але не для всіх еталонних значень вихідного сигналу $y_j^* \in \{0, 1\}$, то перехід до п. 3.

Якщо вагові коефіцієнти w_{ij}, $i \in I_j$, обчислені для всіх комбінацій вхідних сигналів та для всіх еталонних значень вихідного сигналу $y_j^* \in \{0, 1\}$, але не для всіх нейронів N_j, $j \in \{n - n_1 - 1, \ldots, n\}$, то перехід до п. 1.

Якщо вагові коефіцієнти w_{ij}, $i \in I_j$, обчислені для всіх нейронів N_j, $j \in \{n - n_1 - 1, \ldots, n\}$, завершується робота алгоритму.

Кількість ітерацій, потрібних для обчислення вагових коефіцієнтів зв'язків НМ, визначається кількістю нейронів $(n - n_1)$, що не належать першому (вхідному) шару.

Запропонований алгоритм обчислення вагових коефіцієнтів міжнейронних зв'язків НМ працює точно, а його часова складність становить $O(\nu)$ та залежить від розмірності вектора вагових коефіцієнтів w, де ν — кількість зв'язків НМ, $\nu = \sum_{k=1}^{r-1} n_k n_{k+1}$. Збіжність наведеного алгоритму обумовлена скінченністю множини елементів НМ та відсутністю функціональних залежностей між ваговими коефіцієнтами її зв'язків.

Оскільки більшість операцій, передбачених алгоритмом, є операціями привласнення, а операції порівняння використовують тільки для реалізації формули (5) одноразово для кожного зв'язку НМ, можна припустити, що обчислювальна складність алгоритму $O(\nu)$ близька до лінійної.

ПРАКТИЧНА РЕАЛІЗАЦІЯ МЕТОДУ

Практична реалізація запропонованого методу ілюструється прикладом обчислення вагових коефіцієнтів бінарної мережі, призначеної для розв'язання задачі кластеризації.

Припустимо, що характеристики об'єктів кластеризації з урахуванням додаткового вхідного сигналу x_0 описуються чотиривимірними бінарними векторами $x = (x_0, x_1, x_2, x_3)$, $x_i \in \{0, 1\}$, $i = \overline{1, 3}$.

Передбачено три класи об'єктів, номери яких відповідають значенням елементів двовимірного бінарного вектора вихідних сигналів $y = (y_4, y_5)$, $y_j \in \{0, 1\}$, $j = 4, 5$, а саме $y = (1, 0)$ — перший клас, $y = (0, 1)$ — другий клас і $y = (1, 1)$ — третій клас.

Припустимо, що об'єкти, характеристики яких описуються вхідними векторами (1,0,0,0), (0,0,0,1) та (0,0,1,0), належать першому класу; векторами (0,0,1,1), (0,1,0,0) та (0,1,0,1) — другому класу; векторами (0,1,1,0) та (0,1,1,1) — третьому класу.

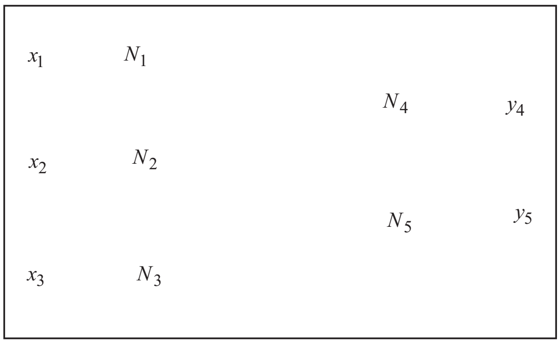

Рис. 2. НМ для розв'язання задачі кластерізації (приклад)

Потрібно обчислити вагові коефіцієнти зв'язків НМ, призначеної для розв'язання цієї задачі кластеризації у наведеній постановці. Для цього формується НМ, що складається з двох шарів і має три вхідні та два вихідні канали (рис. 2).

На цій мережі можна виокремити два елементарні модуля. Перший з них складається з вихідного нейрона N_4 та нейронів вхідного шару N_1, N_2 і N_3, другий — з вихідного нейрона N_5 та тих самих вхідних нейронів N_1, N_2 і N_3.

Алгоритм обчислення вагових коефіцієнтів нейронних зв'язків цієї НМ передбачає послідовне виконання таких дій.

1. Фіксація номера нейрона, який на цій ітерації буде розглядатися як вихідний нейрон елементарного модуля, $j = 5$.

2. Формування множини I_5 номерів нейронів, які безпосередньо передують нейрону N_5, $I_5 = \{1, 2, 3\}$.

3. Вибір еталонного значення вихідного сигналу нейрона N_5 з множини $\{0, 1\}$, $y_5^* = 0$.

4. Формування множин $M_5^{(0)}$ векторів $x_i = (x_0, x_{i5} \mid i \in I_5)$ значень вхідних сигналів нейрона N_5, за яких він повинен видавати еталонний вихідний сигнал $y_5^* = 0$:

$$M_5^{(0)} = \{(1, 0, 0, 0); (0, 0, 0, 1); (0, 0, 1, 0)\}.$$

5. Обчислення вагових коефіцієнтів w_{15}, $i \in I_5$ вхідних зв'язків нейрона N_5 за формулою (5), $w_{15} = w_{25} = w_{35} = 0$.

Виконання відповідно п. 4 та п. 5 алгоритму для еталонного значення вихідного сигналу $y_5^* = 1$:

— формування множин $M_5^{(1)}$ векторів $x_i = (x_0, x_{i5} \mid i \in I_5)$ значень вхідних сигналів нейрона N_5, за яких він повинен видавати еталонний вихідний сигнал $y_5^* = 1$,

$$M_5^{(1)} = \{(0, 0, 1, 1); (0, 1, 0, 0); (0, 1, 0, 1); (0, 1, 1, 0); (0, 1, 1, 1)\};$$

— обчислення вагових коефіцієнтів w_{15}, $i \in I_5$ вхідних зв'язків нейрона N_5 за формулою (5), $w_{15} = w_{25} = w_{35} = 1$.

Виконання відповідно п. 2–5 алгоритму для вихідного нейрона N_4:

— формування множини I_4 номерів нейронів, які безпосередньо передують нейрону N_4, $I_4 = \{1, 2, 3\}$;

— вибір еталонного значення вихідного сигналу нейрона N_4 з множини $\{0, 1\}$, $y_4^* = 0$;

— формування множин $M_4^{(0)}$ векторів $x_i = (x_0, x_{i4} \mid i \in I_4)$ значень вхідних сигналів нейрона N_4, за яких він повинен видавати еталонний вихідний сигнал $y_4^* = 0$,

$$M_4^{(0)} = \{(0, 0, 1, 1); (0, 1, 0, 0); (0, 1, 0, 1)\};$$

— обчислення вагових коефіцієнтів w_{14}, $i \in I_4$ вхідних зв'язків нейрона N_4 за формулою (5), $w_{14} = w_{24} = w_{34} = 0$.

Виконання відповідно п. 4 та п. 5 алгоритму для еталонного значення вихідного сигналу $y_4^* = 1$:

— формування множин $M_4^{(1)}$ векторів $x_i = (x_0, x_{i4} \mid i \in I_4)$ значень вхідних сигналів нейрона N_4, за яких він повинен видавати еталонний вихідний сигнал $y_4^* = 1$,

$$M_4^{(1)} = \{(1, 0, 0, 0); (0, 0, 0, 1); (0, 0, 1, 0); (0, 1, 1, 0); (0, 1, 1, 1)\};$$

— обчислення вагових коефіцієнтів w_{14}, $i \in I_4$ вхідних зв'язків нейрона N_4 за формулою (5), $w_{14} = w_{24} = w_{34} = 1$.

На цьому процес обчислення завершується.

Номер класу об'єкта класифікації визначається двійковим кодом вихідного сигналу.

Обчислення вагових коефіцієнтів вимагає виконання чотирьох ($2n_r = 4$) ітерацій алгоритму з гарантією абсолютної точності результатів розв'язання задачі $\varepsilon = 0$.

Для порівняння запропонованого методу з існуючими підходами задачу було розв'язано за допомогою функції Neuralnet відомої мови машинного навчання R. Для цього множину можливих векторів вхідних сигналів було розбито на три класи за кількістю одиничних елементів. До першого класу увійшли вектори $(1,0,0,0)$, $(0,0,0,1)$, $(0,0,1,0)$ та $(0,1,0,0)$; до другого — вектори $(0,0,1,1)$, $(0,1,0,1)$ та $(0,1,1,0)$; до третього — вектор $(0,1,1,1)$. Кожному класу поставлено у відповідність вихідний сигнал $y_j \in \{0,1\}$, $j = \overline{1,3}$, одиничне значення якого вказує на належність вхідного об'єкта тому чи іншому класу.

Було отримано такі результати обчислення вагових коефіцієнтів: $w_{01} = -6.05332$; $w_{02} = 3.74006$; $w_{03} = 0.28997$; $w_{11} = -10.83074$; $w_{12} = 5.82474$; $w_{13} = 5.46058$; $w_{21} = -6.90986$; $w_{22} = -7.97071$; $w_{23} = 5.2247$; $w_{31} = 7.55311$; $w_{32} = 3.56792$; $w_{33} = -0.45662$.

Середньоквадратична похибка класифікації $\varepsilon = 0.543122$.

Кількість циклів налаштування мережі $L = 106$.

Порівняння отриманих результатів демонструє перевагу запропонованого методу апріорного обчислення вагових коефіцієнтів бінарних НМ перед традиційними методами їхнього навчання для багатьох прикладних областей.

ВИСНОВКИ

Задача обчислення вагових коефіцієнтів НМ є багатоваріантною і має скінченну (у разі дискретних значень сигналів) або нескінченну (коли сигнали вимірюються довільними дійсними числами) множину допустимих розв'язків.

Така властивість задачі пояснюється взаємною незалежністю вагових коефіцієнтів зв'язків, що йдуть до входів нейронів одного шару від нейронів попереднього шару мережі. Це дає змогу декомпозувати загальну задачу обчислення вагових коефіцієнтів НМ на низку підзадач, розв'язуваних окремо для елементарних фрагментів (модулів), що складаються з єдиного вихідного нейрона та множини нейронів попереднього шару, що розглядається як вхідний шар модуля.

Застосування запропонованого методу дає можливість обчислити вагові коефіцієнти НМ на етапі її проєктування та уникнути трудомісткої ітераційної процедури навчання в процесі експлуатації. До того ж, досягається абсолютна відповідність вихідних сигналів НМ їхнім еталонним значенням для всіх заздалегідь передбачених вхідних сигналів.

Метод реалізований для бінарних НМ, але може бути поширений і на мережі інших класів. Наприклад, коли функції активації мають складнішу структуру з довільним зміщенням, а сигнали вимірюються дійсними числами. У цьому випадку вираз (5) набуває іншого вигляду.

Існуючі методи навчання штучних НМ зводяться до реалізації евристичних алгоритмів, що мають увесь комплекс недоліків, властивих алгоритмам цього класу, а саме: відсутність гарантій відшукання розв'язку задачі, коли він об'єктивно існує; висока ймовірність попадання пошукового процесу в тупикові ситуації, що потребують втручання людини (особливо у задачах, постановка яких вимагає дотримання умов у вигляді рівностей); мізерна ймовірність відшукання оптимального розв'язання задачі та ін.

Зрозуміло, що існує безліч задач, розв'язуваних НМ, які можна замінити досить простими Булевими, продукційними або логічними моделями [17], які не потребують трудомістких ітераційних процедур синтезу структури та навчання мережі, але приводять до шуканого результату з абсолютною точністю.

Рішення щодо використання тієї чи іншої моделі залежить від предметної області, що розглядається, і вимагає експертної оцінки.

СПИСОК ЛІТЕРАТУРИ

1. Люгер Дж.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем. Москва: Изд. дом «Вильямс», 2005. 864 с.
2. Рассел С., Норвиг П. Искусственный интеллект: современный подход. Москва: Изд. дом «Вильямс», 2006. 1408 с.
3. Хайкин С. Нейронные сети: полный курс. Москва: Изд. дом «Вильямс», 2006. 1104 с.
4. Каллан Р. Основные концепции нейронных сетей. Москва: Изд. дом «Вильямс, 2001. 288 с.
5. Сигеру О., Марзуки Х., Рубия Ю. Нейроуправление и его приложения. Москва: ИПРЖР, 2000. 272 с.
6. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. Москва: ДМК-Пресс, 2017. 652 с.
7. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: Data mining, inference and prediction. Springer-Verlag, 2009. 746 p.
8. Опанасенко В.Н., Крывый С.Л. Синтез нейроподобных сетей на основе преобразований циклических кодов Хемминга. *Кибернетика и системный анализ.* 2017. Т. 53, № 4. С. 155–164.
9. Палагин А.В., Опанасенко В.Н., Крывый С.Л. Аппаратная реализация преобразований циклического кода Хемминга на базе FPGA. *Труды 15-й Всероссийской научно-технической конференции «Нейроинформатика-2013».* Ч. 3. Москва, 2013. С. 203–212.
10. Kazantsev A.V. Visual data processing and action control using binary neural network. *Eight International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'07), (Santorini, Greece, 6–8 June 2007).* 2007. P. 1–3. https://doi.org//10.1109/WIAMIS.2007.90.
11. Liang S., Sun R., Li Y., Srikant R. Understanding the loss surface of neural networks for binary classification. *Proc. of the 35th International Conference on Machine Learning.* 2018. P. 2835–2843. URL: https://proceedings.mlr.press/v80/liang18a/liang18a.pdf.
12. Krinitskiy M., Verezemskaya P., Grashchenkov K., Tilinina N., Gulev S., Lazzara M. Deep convolutional neural networks capabilities for binary classification of polar mesocyclones in satellite mosaics. *Atmosphere.* 2018. Vol. 9, N 426. P. 1–23.
13. Dunnmon J.A., Yi D., Langlotz C.P., Ré C., Rubin D.L., Lungren M.P. Assessment of convolutional neural networks for automated classification of chest radiographs. *Radiology.* 2018. Vol. 290, N 2. P. 537–544. https://doi.org//10.1148/radiol.2018181422.
14. Korolev S., Safiullin A., Belyaev M., Dodonova Y. Residual and plain convolutional neural networks for 3d brain mri classification. arXiv:1701.06643v1 [cs.CV] 23 Jan 2017. URL: https://arxiv.org/pdf/1701.06643.pdf.
15. Menon A.K., Williamson R.C., The cost of fairness in binary classification. *Proc. of Machine Learning Research.* 2018. Vol. 81. P. 1–12.
16. Ferreyra-Ramírez A., Rodríguez-Martínez E., Avilés-Cruz C., Lypez-Saca F. Image retrieval system based on a binary auto-encoder and a convolutional neural network. *IEEE Latin America Transactions.* 2020. Vol. 18, Iss. 11. P. 1925–1932. https://doi.org//10.1109/TLA.2020.9398634.
17. Litvinenko A. Algorithms for solution inference based on unified logical control models. *Cybernetics and Systems Analysis.* 2020. Vol. 56, N 2. P. 187–194.

A. Litvinenko, D. Kucherov, M. Glybovets

DECOMPOSITION METHOD OF CALCULATING THE WEIGHT COEFFITIENS OF A BINARY NEURAL NETWORK

Abstract. A method for determining the weights of a binary neural network based on its decomposition into elementary modules is presented. The approach allows tuning the weight coefficients of all the network connections at the stage of its design, which has eliminated the need to implement time-consuming iterative algorithms for learning the network during its operation. An algorithm and an example of calculating the weights are given.

Keywords: binary neural network, weights, method of determination, decomposition, algorithm.

Надійшла до редакції 26.06.2022