УДК 004.8: 004.032.26

Є.В. Бодянський [1] , Т.Є. Антоненко [2]

[1] Професор кафедри штучного інтелекту, ХНУРЕ, м. Харків, Україна, yevgeniy.bodyanskiy@nure.ua;

[2] Аспірант, ХНУРЕ, м. Харків, Україна, tymofii.antonenko@nure.ua

ГЛИБОКА НЕО-ФАЗЗІ НЕЙРОННА МЕРЕЖА ТА ЇЇ НАВЧАННЯ

Оптимізація швидкодії навчання глибоких нейронних мереж є надзвичайно актуальним питанням. Сучасні підходи орієнтуються на використання нейронних мереж на основі персептрону Розенблатта. Але отримувані результати не являються задовільними для індустріальних та наукових потреб в контексті швидкодії навчання нейронних мереж. Також такий підхід натикається на проблеми зникаючого та вибухаючого градієнта. Для вирішення проблеми в статті запропоновано використовувати нео-фаззі нейрон, властивості якого основані на F-перетворенні. В статті розглянуто використання нео-фаззі нейрона як основного компонента нейронної мережі. Показана архітектура глибокої нео-фаззі нейронної мережі а також алгоритм зворотнього поширення похибки для цієї архітектури з трикутною функцією принадлежності для нео-фаззі нейрона. Приведені основні переваги щодо застосування нео-фаззі нейрона як основного компоненту нейронної мережі. В статті описано за рахунок яких властивостей нео-фаззі нейрона вирішуються питання покращення швидкодії та зникаючого чи вибухаючого градієнта. Порівняно запровоновану архітектуру нео-фаззі глибокої нейронної мережі зі стандартними глибокими мережами на основі персептрону Розенблатта.

НЕО-ФАЗЗІ НЕЙРОН, БАГАТОШАРОВА НЕЙРОННА МЕРЕЖА, F-ПЕРЕТВОРЕННЯ

Е.В. Бодянский, Т.Е. Антоненко. Глубокая нео-фаззи нейронная сеть и ее обучение. Оптимизация быстродействия обучения глубоких нейронных сетей является чрезвычайно актуальным вопросом. Современные подходы ориентируются на использование нейронных сетей на основе персептрона Розенблатта. Но получаемые результаты не являются удовлетворительными для индустриальных и научных потребностей в контексте быстродействия обучения нейронных сетей. Также такой подход натыкается на проблемы исчезающего и взрывающегося градиента. Для решения проблемы в статье предложено использовать нео-фаззи нейрон, свойства которого основаны на F-преобразовании. В статье рассмотрено использование нео-фаззи нейрона как основного компонента нейронной сети. Показана архитектура глубокой нео-фаззи нейронной сети, а также алгоритм обратного распространения ошибки для этой архитектуры с треугольной функцией принадлежности для нео-фаззи нейрона. Приведены основные преимущества по применению нео-фаззи нейрона как основного компонента нейронной сети. В статье описано за счет каких свойств нео-фаззи нейрона решаются вопросы улучшения быстродействия и исчезающего или взрывающейся градиента. Проведено сравнение предложенной архитектуры нео-фаззи глубокой нейронной сети со стандартными глубокими сетями на основе персептрона Розенблатта.

НЕО-ФАЗЗИ НЕЙРОН, МНОГОСЛОЙНАЯ НЕЙРОННАЯ СЕТЬ, F-ПРЕОБРАЗОВАНИЕ

Ye. Bodyankiy. T. Antonenko. Deep neo-fuzzy neural network and its learning. Optimizing the learning speed of deep neural networks is an extremely important issue. Modern approaches focus on the use of neural networks based on the Rosenblatt perceptron. But the results obtained are not satisfactory for industrial and scientific needs in the context of the speed of learning neural networks. Also, this approach stumbles upon the problems of a vanishing and exploding gradient. To solve the problem, the paper proposed using a neo-fuzzy neuron, whose properties are based on the F-transform. The article discusses the use of neo-fuzzy neuron as the main component of the neural network. The architecture of a deep neo-fuzzy neural network is shown, as well as a backpropagation algorithm for this architecture with a triangular membership function for neo-fuzzy neuron. The main advantages of using neo-fuzzy neuron as the main component of the neural network are given. The article describes the properties of a neo-fuzzy neuron that addresses the issues of improving speed and vanishing or exploding gradient. The proposed neo-fuzzy deep neural network architecture is compared with standard deep networks based on the Rosenblatt perceptron.

NEO-FUZZY NEURON, MULTILAYER NEURAL NETWORK, F-TRANSFORM

Вступ

В цей час штучні нейронні мережі (ШНМ) отримали широке розповсюдження для вирішення задач опрацювання інформації різної природи завдяки своїм універсальним апроксимуючим властивостям що досягаються в процесі їх навчання на основі існуючих даних спостережень. Більш високих результатів можливо досягти за допомогою глибоких нейронних мереж (ГНМ) [1-5], котрі хоча й перевершують за якістю рішення задач традиційні нейронні мережі, але їх навчання пов'язано з низкою суттєвих обчислювальних проблем та вимагає більших витрат часу. В основі більшості ШНМ лежить, так званий, елементарний персептрон Розенблатта [6], при цьому як функція активації використовується традиційна сигмоїда або гіперболічний тангенс. Трьохшаровий персептрон забезпечує високу якість апроксимації достатньо складних функцій, заданих в обмеженій області визначення [8]. Спроби покращити якість отриманого рішення шляхом збільшення кількості прихованих шарів ШНМ натикаються на проблему, пов'язану з, так званим, зникаючим та вибуховим градієнтом [9], поява котрого пов'язана з формою

сигмоїдальних активаційних функцій. У зв'язку з цим ГНМ замість сигмоїдальних функцій зазвичай використовують *linear rectified family*, характерним представником котрого є *rectified linear unit* (ReLU), який не страждає від проблем, пов'язаних з обчисленням градієнта. Використання таких функцій не викликає проблем, пов'язаних з обчисленням градієнта, а їх похідні-константи дозволяють спростити дельта-правило навчання кожного окремого нейрона. Разом з тим, використання функції *ReLU* не відповідає вимогам апроксимаційних теорем, які лежать в основі ШНМ та, поперед всього, вимогам теореми Цибенко [7]. Таким чином, ГНМ забезпечують кусково-лінійну апроксимацію, висока якість котрої досягається збільшенням числа прихованих шарів мережі. При цьому збільшення кількості шарів зменшує швидкодію процесу навчання та вимагає більших об'ємів даних для навчання.

1. Архітектура глибокої нео-фаззі нейронної мережі

Нейронна мережа, яку ми розглядаємо, має традиційну багатошарову архітектуру прямого поширення, включаючи в загальному випадку *s* шарів опрацювання інформації.

На вхідний (нульовий) шар подається $x(k) \in R^n$ вектор вхідних сигналів.

$$x(k) = (x_1(k), x_2(k), \ldots, x_n(k)),$$

де k=1, 2, ..., N – номер спостереження в навчальній виборці або індекс поточного дискретного часу. Вихідним сигналом мережі є вектор

$$\hat{y}(k) = (\hat{y}_1(k), \hat{y}_2(k), \ldots, \hat{y}_m(k))^T \in R^m.$$

Далі, для спрощення запису позначень будемо також використовувати вид

$$x(k) \equiv o^{[0]}(k) = (o_1^{[0]}(k), \ldots, o_{i_0}^{[0]}(k) \ldots, o_{n_0}^{[0]}(k))^T,$$

$$\hat{y}(k) \equiv o^{[s]}(k) = (o_1^{[s]}(k), \ldots, o_{i_s}^{[s]}(k), \ldots, o_{n_s}^{[s]}(k))^T.$$

Таким чином, вхідним сигналом p-го шару (p=1, 2..., s) є вектор

$$o^{[p-1]}(k) = (o_1^{[p-1]}(k), \ldots, o_{i_{p-1}}^{[p-1]}(k), \ldots, o_{n_{p-1}}^{[p-1]}(k))^T \in R^{n_{p-1}},$$

а вихідним - вектор

$$o^{[p]}(k) = (o_1^{[p]}(k), \ldots, o_{i_p}^{[p]}(k), \ldots, o_{n_p}^{[p]}(k))^T \in R^{n_p}.$$

При цьому нео-фаззі нейронна мережа містить $\sum_{p=1}^{s} n_p$ нейронів.

Вузлом цієї архітектури є нео-фаззі нейрон (НФН) [10-12] з n_{p-1} входами та одним виходом $o_{i_p}^{[p]}$. На рис. 1 зображено архітектуру i_p-го $NFN_{i_p}^{[p]}$ p-го шару мережі.

Кожен i_p-ий ($i_p = 1,2,\ldots, n_p$) нео-фаззі нейрон p-го ($p = 1,2,\ldots, s$) шару нео-фаззі нейронної мережі містить n_{p-1} нелінійних синапсів $NS_{i_p i_{p-1}}^{[p]}$, кожен з котрих включає h функцій належності $\mu_{i_p i_{p-1} l}^{[p]}$ ($l = 1,2,\ldots, h$) і таку ж кількість синаптичних вагових коефіцієнтів $w_{i_p i_{p-1} l}^{[p]}$, які налаштовуються в процесі навчання. Таким чином, ця архітектура має $\sum_{p=1}^{s} n_p n_{p-1}$ нелінійних синапсів та $h\sum_{p=1}^{s} n_p n_{p-1}$ функцій належності $\mu_{i_p i_{p-1} l}^{[p]}(o_{i_{p-1}}^{[p-1]})$ та таку ж кількість налаштованих синаптичних вагових коефіцієнтів $w_{i_p i_{p-1} l}^{[p]}$.

Вихідний сигнал кожного нелінійного синапсу $NS_{i_p i_{p-1}}^{[p]}$ може бути записаний у вигляді

$$f_{i_p i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = \sum_{l=1}^{h} w_{i_p i_{p-1} l}^{[p]} \mu_{i_p i_{p-1} l}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right), \quad (1)$$

а вихідний сигнал нео-фаззі нейрону $NFN_{i_p}^{[p]}$ в цілому

$$o_{i_p}^{[p]} = \sum_{i_{p-1}=1}^{n_{p-1}} f_{i_p i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = \sum_{i_{p-1}=1}^{n_{p-1}} \sum_{l=1}^{h} w_{i_p i_{p-1} l}^{[p]} \mu_{i_p i_{p-1} l}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right). \quad (2)$$

Для вихідного шару мережі сигнал (2) можна також записати у формі

$$\hat{y}_j = o_{i_s}^{[s]} = \sum_{i_{s-1}=1}^{n_{s-1}} f_{i_s i_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}\right) = \sum_{i_{s-1}=1}^{n_{s-1}} \sum_{l=1}^{h} w_{i_s i_{s-1} l}^{[s]} \mu_{i_s i_{s-1} l}^{[s]}\left(o_{i_{s-1}}^{[s-1]}\right). \quad (3)$$

Далі, вводячи в розгляд синаптичні ваги та функції належності

$$w_{i_p i_{p-1}}^{[p]} = (w_{i_p i_{p-1} 1}^{[p]}, \ldots, w_{i_p i_{p-1} l}^{[p]}, \ldots w_{i_p i_{p-1} h}^{[p]})^T,$$

$$\mu_{i_p i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) =$$

$$= (\mu_{i_p i_{p-1} 1}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right), \ldots \mu_{i_p i_{p-1} l}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right), \ldots \mu_{i_p i_{p-1} h}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right))^T$$

розмірності ($h \times 1$), а також

$$w_{i_p}^{[p]} = (w_{i_p 1}^{[p]}, \ldots, w_{i_p i_{p-1}}^{[p]}, \ldots w_{i_p n_{p-1}}^{[p]})^T$$

та

$$\mu_{i_p}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) =$$

$$= (\mu_{i_p 1}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right), \ldots \mu_{i_p i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right), \ldots \mu_{i_p n_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right))^T$$

розмірності ($n_p h \times 1$), можна записати

$$f_{i_p i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = w_{i_p}^{[p]T} \mu_{i_p}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) \text{ замість (1),}$$

$$o_{i_p}^{[p]} = w_{i_p}^{[p]T} \mu_{i_p}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) \text{ замість (2)}$$

$$\text{і } \hat{y}_j = o_{i_s}^{[s]} = w_{i_s}^{[s]T} \mu_{i_s}^{[s]}\left(o_{i_{s-1}}^{[s-1]}\right) \text{ замість (3).}$$

В якості функції належності нелінійних сигналів $NS_{i_p i_{p-1}}^{[p]}$ автори нео-фаззі нейрона [10-12] використовували традиційну трикутну функцію, яка задовольняє вимогам одиничного розбиття Руспіні:

$$\mu_{i_p i_{p-1} l}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) =$$

$$= \begin{cases} \dfrac{o_{i_{p-1}}^{[p-1]} - c_{i_p i_{p-1} l-1}^{[p]}}{c_{i_p i_{p-1} l}^{[p]} - c_{i_p i_{p-1} l-1}^{[p]}}, \text{ if } o_{i_{p-1}}^{[p-1]} \in [c_{i_p i_{p-1} l-1}^{[p]}, c_{i_p i_{p-1} l}^{[p]}) \\ \dfrac{c_{i_p i_{p-1} l+1}^{[p]} - o_{i_{p-1}}^{[p-1]}}{c_{i_p i_{p-1} l+1}^{[p]} - c_{i_p i_{p-1} l}^{[p]}}, \text{ if } o_{i_{p-1}}^{[p-1]} \in [c_{i_p i_{p-1} l}^{[p]}, c_{i_p i_{p-1} l+1}^{[p]}) \\ 0, \text{ інакше,} \end{cases} \quad (4)$$

де $c_{i_p i_{p-1} l}^{[p]}, l = 1,2,\ldots,h$ — центри трикутних функцій належності. У випадку, коли всі нелінійні синапси

мережі $NS_{i_p i_{p-1}}^{[p]}$ мають однакове число h центрів, котрі розподілені рівномірно по осям вхідних сигналів, вираз (4) може бути переписаний у більш зручному вигляді:

$$\mu_{i_p i_{p-1} l}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right)= \\ =\begin{cases}\Delta_c^{-1}(o_{i_{p-1}}^{[p-1]}-c_{i_p i_{p-1} l-1}^{[p]}), \text{ if } o_{i_{p-1}}^{[p-1]}\in[c_{i_p i_{p-1} l-1}^{[p]}, c_{i_p i_{p-1} l}^{[p]}) \\ \Delta_c^{-1}(c_{i_p i_{p-1} l+1}^{[p]}-o_{i_{p-1}}^{[p-1]}), \text{ if } o_{i_{p-1}}^{[p-1]}\in[c_{i_p i_{p-1} l}^{[p]}, c_{i_p i_{p-1} l+1}^{[p]}) \\ 0, \text{ інакше,}\end{cases} \quad (5)$$

при цьому:

$$\mu_{i_p i_{p-1} l-1}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right)+\mu_{i_p i_{p-1} l}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right)= \\ \mu_{i_p i_{p-1} l}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right)+\mu_{i_p i_{p-1} l+1}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right)=1. \quad (6)$$

Умови (4)-(6) означають, що в кожен момент часу k в кожному нелінійному синапсі тільки дві сусідні функції належності можуть спрацювати. Це приводить до того, що в цей же момент налаштовуються тільки два сусідні синаптичні вагові коефіцієнти, тобто на кожному кроці навчання уточнюється тільки $2\sum_{p=1}^{s} n_{p-1} n_p$ синаптичних ваг замість

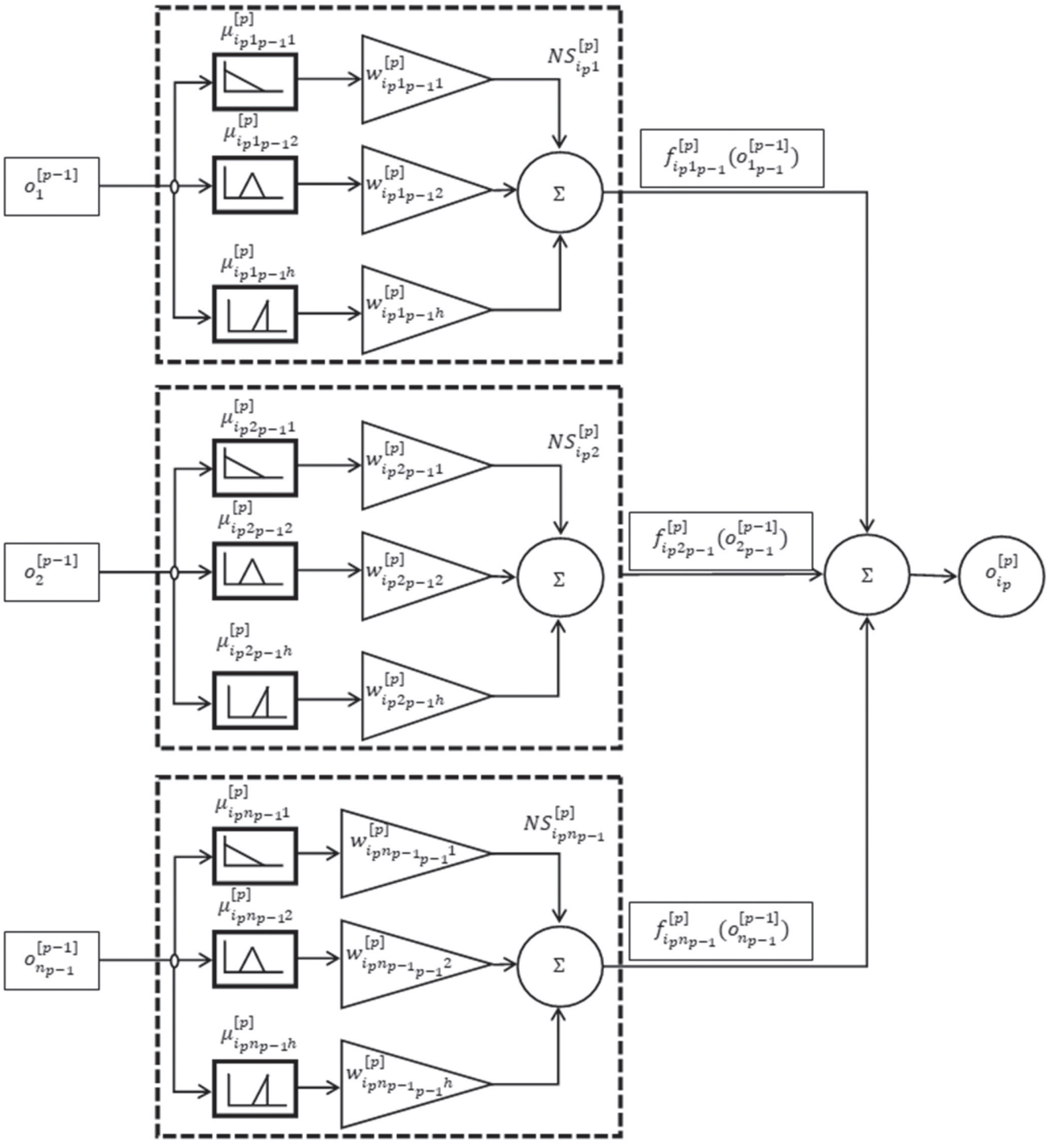

Рис. 1. *i*-ий нео-фаззі нейрон *p*-го шару нео-фаззі нейронної мережі

$h\sum_{p=1}^{s} n_{p-1} n_p$. Оскільки кожен нелінійний синапс мережі реалізує по суті F-перетворення [17], це дозволяє як завгодно точно апроксимувати будь-яку обмежену одновимірну функцію при достатньо великому h. Збільшення числа функцій належності не ускладнює навчання мережі.

Таким чином, кожен NFN є гібридом нео-фаззі системи Ванга-Менделя та F-перетворення, що забезпечує високі апроксимаційні властивості мережі в цілому.

2. Навчання глибокої нео-фаззі нейронної мережі

Навчання багатошарової нео-фаззі мережі реалізується на основі зворотнього поширення похибки.

Процес навчання нео-фаззі мережі зводиться до пошуку синаптичних вагових коефіцієнтів

$$w^{[p]}_{i_p i_{p-1} l}, p = 1,2,\ldots,s; l = 1,2,\ldots,h;$$
$$i_p = 1,2,\ldots,n_p; i_{p-1} = 1,2,\ldots,n_{p-1}$$

шляхом мінімізації прийнятої цільової функції навчання.

В якості критерію навчання використаємо стандартну квадратичну функцію

$$E(k) = \frac{1}{2}\sum_{i_s=1}^{n_s} e^2_{i_s}(k) = \frac{1}{2}\sum_{j=1}^{m} e^2_j(k), \tag{7}$$

де $e^2_j(k) = (y_j(k) - w^{[s]}_{i_s}(k-1)\mu^{[s]}_{i_s}(o^{[s-1]}_{i_{s-1}}(k)))^2$,

$y_j(k)$ – зовнішній навчальний сигнал.

Градієнтна мінімізація (7) для кожного синаптичного вагового коефіцієнту вихідного шару $w^{[s]}_{i_s i_{s-1} l}$ може бути описана рекурентною процедурою

$$\begin{aligned} w^{[s]}_{i_s i_{s-1} l}(k) &= w^{[s]}_{i_s i_{s-1} l}(k-1) - \eta(k)\frac{\partial e^2_{i_s}(k)}{\partial w^{[s]}_{i_s i_{s-1} l}} = \\ &= w^{[s]}_{i_s i_{s-1} l}(k-1) - \eta(k)\mu^{[s]}_{i_s}(O^{[s-1]}_{i_{s-1}}(k)) \end{aligned} \tag{8}$$

(тут $\eta(k)$ – параметр кроку навчання), або у векторній формі

$$\begin{aligned} w^{[s]}_{i_s}(k) = w^{[s]}_{i_s}(k-1) - \eta(k)(y_j(k) - \\ - w^{[s]}_{i_s}(k-1)\mu^{[s]}_{i_s}\left(o^{[s-1]}_{i_{s-1}}(k)\right). \end{aligned} \tag{9}$$

Для вибору $\eta(k)$ може бути використана процедура, яка є гібридом оптимального алгоритму Kaczmarz-Widrow-Hoff та стохастичної апроксимації [18, 19].

$$\begin{cases} w^{[s]}_{i_s}(k) = w^{[s]}_{i_s}(k-1) - r^{-1}_{i_s}(k) e^2_{i_s}(k)\mu^{[s]}_{i_s}\left(o^{[s-1]}_{i_{s-1}}(k)\right), \\ r^{-1}_{i_s}(k) = \alpha r_{i_s}(k-1) + \mu^{[s]}_{i_s}(o^{[s-1]}_{i_{s-1}}(k))^2, \end{cases}$$

де $0 \le \alpha \le 1$ – фактор забування.

Налаштування синаптичних вагових коефіцієнтів прихованих шарів реалізується за допомогою алгоритму зворотного поширення похибки, для чого може бути використана процедура типу (8) у вигляді

$$\begin{aligned} w^{[s-1]}_{i_{s-1} i_{s-2} l}(k) &= w^{[s-1]}_{i_{s-1} i_{s-2} l}(k-1) - \eta(k)\frac{\partial E(k)}{\partial w^{[s-1]}_{i_{s-1} i_{s-2} l}} = \\ &= w^{[s-1]}_{i_{s-1} i_{s-2} l}(k-1) - \eta(k)\frac{\partial e^2_{i_s}(k)}{\partial w^{[s]}_{i_s i_{s-1} l}}\frac{\partial o^{[s]}_{i_s}(k)}{\partial o^{[s-1]}_{i_{s-1}}(k)}\frac{\partial o^{[s-1]}_{i_{s-1}}(k)}{\partial w^{[s]}_{i_{s-1} i_{s-2} l}}. \end{aligned} \tag{10}$$

Оскільки

$$\frac{\partial o^{[s]}_{i_s}}{\partial o^{[s-1]}_{i_{s-1}}} = \frac{\partial f^{[s]}_{i_s i_{s-1} l}(o^{[s-1]}_{i_{s-1}})}{\partial o^{[s-1]}_{i_{s-1}}} = \sum_{l=1}^{h} w^{[s]}_{i_s i_{s-1} l}\frac{\partial \mu^{[s]}_{i_s i_{s-1} l}(o^{[s-1]}_{i_{s-1}})}{\partial o^{[s-1]}_{i_{s-1}}},$$

де

$$\frac{\partial \mu^{[s]}_{i_s i_{s-1} l}(o^{[s-1]}_{i_{s-1}})}{\partial o^{[s-1]}_{i_{s-1}}} = \begin{cases} \Delta_c^{-1}, & \text{if } o^{[p-1]}_{i_{p-1}} \in \left[c^{[p]}_{i_p i_{p-1} l-1}, c^{[p]}_{i_p i_{p-1} l}\right), \\ -\Delta_c^{-1}, & \text{if } o^{[p-1]}_{i_{p-1}} \in \left[c^{[p]}_{i_p i_{p-1} l}, c^{[p]}_{i_p i_{p-1} l+1}\right), \\ 0, & \text{інакше}, \end{cases} \tag{11}$$

алгоритм (10) остаточно може бути записаний у формі

$$\begin{aligned} w^{[s-1]}_{i_{s-1} i_{s-2} l}(k) &= w^{[s-1]}_{i_{s-1} i_{s-2} l}(k-1) - \\ &- \eta(k)\sum_{i_s=1}^{n_s}\partial e^2_{i_s}(k)\sum_{l=1}^{h} w^{[s]}_{i_s i_{s-1} l}(k) * \\ &* \frac{\partial \mu^{[s]}_{i_s i_{s-1} l}\left(o^{[s-1]}_{i_{s-1}}\right)}{\partial o^{[s-1]}_{i_{s-1}}}\mu^{[s-1]}_{i_{s-1} i_{s-2} l}(o^{[s-2]}_{i_{s-2}}(k)) \end{aligned} \tag{12}$$

Аналогічно (12) може бути записаний алгоритм налаштування для p-го нейронного шару, при чому похибка нео-фаззі нейрона p-го шару має вигляди:

$$\frac{\partial E}{\partial o^{[p]}_{i_p}} = \sum_{i=1}^{n_{p+1}}\frac{\partial E}{\partial o^{[p+1]}_{i_{p+1}}}\frac{\partial o^{[p+1]}_{i_{p+1}}}{\partial o^{[p]}_{i_p}},$$

де дельта-похибка має вигляд

$$\delta^{[p]}_j = \frac{\partial E}{\partial o^{[p]}_j} = \sum_{i=1}^{n_{p+1}}\delta^{[p+1]}_j\frac{\partial o^{[p+1]}_{i_{p+1}}}{\partial o^{[p]}_{i_p}} \tag{13}$$

Тоді для всіх вагових коефіцієнтів можна записати

$$\frac{\partial E(k)}{\partial w^{[p]}_{i_p i_{p-1} l}} = \sum_{i=1}^{n_{p+1}}\delta^{[p+1]}_j\frac{\partial o^{[p+1]}_i}{\partial o^{[p]}_j}\frac{\partial o^{[p]}_i}{\partial w^{[p]}_{i_p i_{p-1} l}},$$

де дельта похибки береться з наступного шару, а проміжні похідні обчислюються наступним чином:

$$\frac{\partial o^{[p+1]}_{i_{p+1}}}{\partial o^{[p]}_{i_p}} = \frac{\partial f^{[p+1]}_{i_s i_{s-1} l}(o^{[p]}_{i_p})}{\partial o^{[p]}_{i_p}} = \sum_{l=1}^{h} w^{[p+1]}_{i_{p+1} i_p l}\frac{\partial \mu^{[p+1]}_{i_{p+1} i_p l}(o^{[p]}_{i_p})}{\partial o^{[p]}_{i_p}},$$

де права частина отримується за допомогою (11).

Похідна вихідного сигналу нео-фаззі нейрона по ваговому коефіцієнту має вигляд:

$$\frac{\partial o^{[p]}_i}{\partial w^{[p]}_{i_p i_{p-1} l}} = \frac{f^{[p]}_{i_p i_{p-1}}\left(o^{[p]}_i\right)}{\partial w^{[p]}_{i_p i_{p-1} l}} = \mu^{[p]}_{i_p i_{p-1} l}\left(o^{[p-1]}_{i_{p-1}}\right),$$

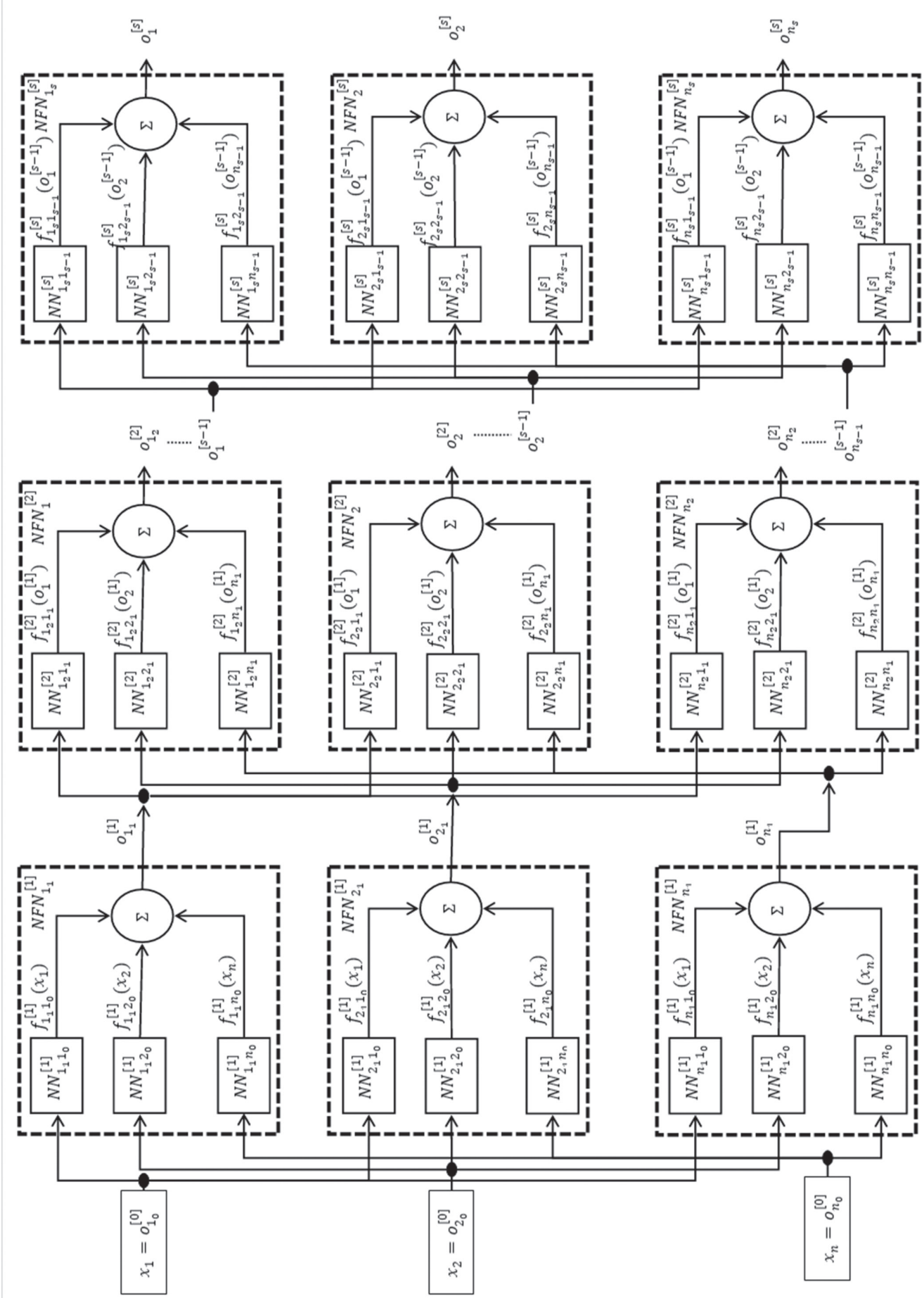

Рис. 2. Архітектура багатошарової нео-фаззі нейронної мережі

де остаточний вираз

$$w^{[p]}_{i_p i_{p-1} l}(k) = w^{[p]}_{i_p i_{p-1} l}(k-1) - \eta(k)\frac{\partial E(k)}{\partial w^{[p]}_{i_p i_{p-1} l}}$$

може бути записано у вигляді:

$$w^{[p]}_{i_p i_{p-1} l}(k) = w^{[p]}_{i_p i_{p-1} l}(k-1) - \\ -\eta(k)\sum_{i=1}^{n_{p+1}}\delta^{[p+1]}_j \frac{\partial o^{[p+1]}_i}{\partial o^{[p]}_j}\mu^{[p]}_{i_p i_{p-1} l}(o^{[p-1]}_{i_{p-1}}) \qquad (14)$$

Нескладно бачити, що навчання нео-фаззі нейронної мережі, архітектура якої в цілому наведена на рис. 2, з обчислювальної точки зору простіше в порівнянні з ШНМ та ГНМ побудованими на основі традиційних персептронів Розенблата, оскільки похідні трикутних функцій належності є константами.

3. Експериментальні дослідження

Порівнювалися між собою результати навчання на якість апроксимації багатошарового персептрону з сигмоїдальними та ReLU-функціями активації та запропонованої нео-фаззі мережі. Кожна модель мала 4 прихованих шари та по 50 нейронів на кожному прихованому шарі. Результати експериментів показали, що нео-фаззі мережа досягає тієї ж якості що й багатошаровий персептрон за меншу кількість епох навчання.

Висновки

Запропоновано архітектуру та алгоритм навчання глибокої нео-фаззі нейронної мережі, основною відмінністю якої від традиційних багатошарових нейронних мереж з прямим поширенням інформації є використання в якості вузлів нео-фаззі нейронів замість традиційних елементарних персептронів Розенблата. Запропонована нео-фаззі нейронна мережа має високі апроксимаційні властивості та характеризується підвищеною швидкістю навчання своїх синаптичних ваг, завдяки використанню трикутних функцій належності в нелінійних синапсах нео-фаззі нейронів. Крім того, запропонована глибока нео-фаззі нейронна мережа є досить простою з точки зору обчислювальної реалізації.

Список літератури:

[1] *Bengio Y, LeCun Y, Hinton G.* Deep Learning – Nature – 2015-521 – p.436-444.

[2] *Schmidhuber J.* Deep learning in neural networks: An overview – Neural Networks – 2015–01 – p.85-117.

[3] *Googfellow I., Bengio Y., Courville A.* Deep Learning – MIT Press, 2016-787p.

[4] *Graupe D.* Deep Learning Neural Networks: Design and Case Studies- New York: World Scientific, 2016 – 260p.

[5] *Caterini A.L., Chang D.E.* Deep Neural Networks in a Mathematical Framework – Springer, 2018 –79p.

[6] *Cichocki A., Unbehauen R.* Neural Networks for Optimization and Signal Processing – Stuttgart: Teubner, 1993-526p.

[7] *Cybenko G.* Approximation by superpositions of a sigmoidal function – Math. Control Signals Systems. – 1985 – 2 – p.303-314.

[8] *Hornik K.* Approximation capabilities of multilayer feedforward networks – 1994 – 4 – p.251-257.

[9] *Aggarwal Ch.C.* Neural Networks and Deep Learning – Cham: Springer, 2018-512p.

[10] *Yamakawa T, Uchino E, Miki T., Kusabagi H.* A neo fuzzy neuron and its applications to system identification and predictions to system behavior. – Proc. 2nd Int. Conf. on Fuzzy Logic and Neural Networks, pp. 477-483, 1992.

[11] *Uchino E, Yamakava T.* Neo-fuzzy neuron based new approach to system modeling with application to actual system - Proceedings Sixth International Conference on Tools with Artificial Intelligence – New Orlean, LA, USA, 1994 – p.564-570.

[12] *Miki T., Yamakawa T.,* "Analog implementation of neo-fuzzy neuron and its on-board learning," In Computational Intelligence and Applications, Piraeus: WSES Press, 1999, pp. 144-149.

[13] *Kolodyazhniy V, Bodyanskiy Ye.* Fuzzy Kolmogorov's network – Lecture Notes in Computer Science. – 3214 – Heidelberg: Springer Verlag, 2004. – p.764-771.

[14] *Bodyanskiy Ye., Kolodyazhniy V., Otto P.* Neuro-fuzzy Kolmogorov's network for time series prediction and pattern classification – Lecture Notes in Artificial Intelligence – 3698 – Heidelberg: Springer Verlog, 2005. – p.191-202.

[15] *Bodyanskiy Ye., Popov S., Rybalchenko T.* Multilayer neuro-fuzzy network for short term electric load forecasting – Lecture Notes in Computer Science. – 5010 – Berlin-Heidelberg: Springer Verlag, 2008. – p.339-348.

[16] *Bodyanskiy Ye., Vynokurova O., Setlak G., Peleshko D., Mulesa P.* Adaptive multivariate hybrid neuro-fuzzy system and its on-board fast learning – Neurocomputing – 2017 – 230-p.409-416.

[17] *Perfilieva T.* Fuzzy transforms: Theory and applications – Fuzzy Sets and Systems – 2006 – 157 – p.993-1023.

[18] *Bodyanskiy Ye., Kolodyazhniy V., Stephan A.* An adaptive learning algorithm for a neuro-fuzzy network – Ed. by B.Reush "Computitional Intelligence. Theory and Application" – Berlin-Heidelberg: Ney York: Springer, 2001. – p.68-75.

[19] *Otto P., Bodyanskiy Ye., Kolodyazhniy V.* A new learning algorithm for a forecasting neuro-fuzzy network - Integrated Computer Aided Engineering – 2003 – 10(4) – p.399-409.

Надійшла до редколегії 10.04.2019