УДК 681.14

Коцовський В. М., Вашкеба М. М.
ДВНЗ «Ужгородський національний університет»

Навчання штучних комплексних нейронних мереж методом зворотного поширення помилки

В роботі розглядаються нейромережі з комплексними вагами. Наведений алгоритм навчання цих нейромереж, який базується на методі зворотного поширення похибки.

Ключові слова: нейрон, комплексний нейрон, нейромережа, алгоритми навчання, алгоритм зворотного поширення.

Neural networks with complex weights have been studied in the paper. Learning algorithms based on backpropagation method has been given in the paper.

Key Words: neuron, complex neuron, neural network, learning algorithms, backpropagation.

1. Вступ

Штучні нейронні мережі (ШНМ) успішно використовуються для розв’язування широкого кола практичних та теоретичних проблем [1-2]. На сьогоднішній час (ШНМ) — ефективний інструмент для розв’язування задач апроксимації функцій, прогнозування поведінки динамічних систем, класифікації множини об’єктів по кільком ознакам, розпізнавання образів, асоціативного пошуку та багатьох інших. У нейроінформатиці розроблено багато підходів, які стосуються архітектури НМ з дійсними ваговими коефіцієнтами, вигляду функцій активацій нейронів (неперервні або розривні порогового типу) та запропоновано цілий ряд алгоритмів навчання НМ. НЕ з дискретними функціями активації вивчалися в [3]. В даній роботі вводиться поняття комплексного нейрона з функцією активації, яка має неперервно диференційовані частинні похідні по дійсній і уявній частині аргументів, розглядаються ШНМ, побудовані з таких елементів, та описується модифікація відомого алгоритму зворотного поширення для навчання комплексних ШНМ. При цьому комплексні НМ можуть бути використані як для розв’язування тих самих задач, що й звичайні НМ (із можливим зменшенням кількості входів і виходів мережі), так і для розв’язування

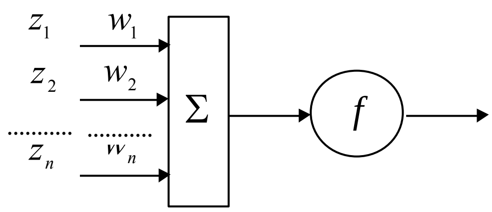

Рис.1. Комплексний нейрон.

задач із специфічними комплексними вхідними і вихідними даними (наприклад наближення функцій комплексної змінної).

2. Основні поняття та означення

Неперервний комплексний нейронний елемент (нейрон з неперервною комплексною функцією активації) (НКНЕ) — функціональний елемент із n входами та одним виходом y, який обчислюється наступним чином:

$$y = f\left(\sum_{j=1}^{n} w_j z_j + w_{n+1}\right),$$

де комплексні числа $z_1,\ldots,z_n$ – вхідні сигнали, $w_1,\ldots,w_n,w_{n+1}$ – вагові комплексні коефіцієнти (коефіцієнт w_{n+1} по аналогії до [3] можна назвати порогом НКНЕ), $f: C \to C$ – нелінійна, неперервна разом із своїми частинними похідними функція, яка називається функцією активації. НКНЕ різними способами можна з’єднувати у комплексну штучну нейронну мережу КШНМ. Ми обмежимося розглядом багатошарових КШНМ прямого поширення, тобто мереж, які задовольняють наступну умову: нейрони кожного шару з’єднуються з нейронами попереднього та наступного шарів по принципу “кожний з кожним”. Перший шар називається вхідним, внутрішні – прихованими, останній – вихідним. Робота КШНМ описується формулами:

$$y_{kl} = f\left(\sum_{j} w_{jkl} z_{jkl} + w_{n+1,kl}\right),\ x_{kj,l+1} = y_{kl},$$

де індекс j позначає номер вхідного нейрона, k – номер вихідного нейрона, l – номер шару, $z_{jkl} = x_{jkl} + i\,y_{jkl}$ – значення j-го вхідного сигналу k-го нейрона в шарі l, $w_{jkl} = u_{jkl} + i\,v_{jkl}$ – значення j-го вагового коефіцієнта k-го нейрона в шарі l.

Багатошарова КШНМ обчислює вихідний вектор $F(\mathbf{z})$ на основі вхідного вектора $\mathbf{z}$. Під алгоритмом навчання КШНМ з учителем будемо розуміти підбір параметрів мережі (вагових коефіцієнтів w_{jkl}) таким чином, щоб КШНМ ставила у відповідність вхідним векторам із множини $\{\mathbf{z}^1,\ldots,\mathbf{z}^m\}$ відповідні їм вихідні вектори з множини $\{\mathbf{d}^1,\dots,\mathbf{d}^m\}$. Сукупність пар $\{(\mathbf{z}^1,\mathbf{d}^1),\dots,(\mathbf{z}^m,\mathbf{d}^m)\}$ називається навчальною множиною. Нехай f_k^t – значення вихідного сигналу *k*-го нейрона в

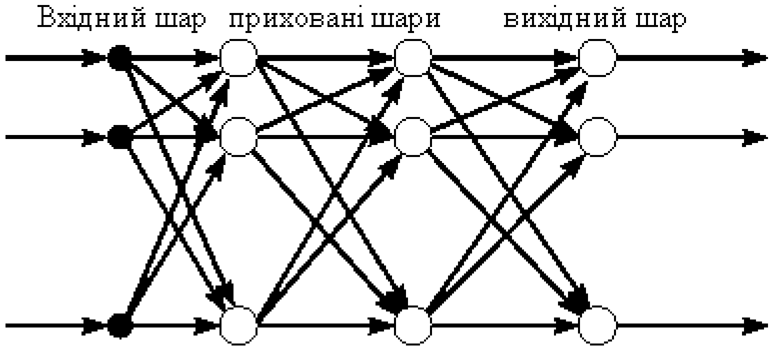

Рис. 2. Багатошарова НМ прямого поширення.

останньому (вихідному) шарі *l* у випадку, коли на вхід НМ подано вектор $\mathbf{z}^t$ Під похибкою мережі будемо розуміти величину

$$E = \frac{1}{2}\sum_k \sum_t \left| f_k^t - d_k^t \right|^2 . \tag{1}$$

Будемо вважати, що $E = E(\mathbf{w}) = E(\mathbf{u},\mathbf{v})$, де $\mathbf{u}$ – вектор, компонентами якого є дійсні частини всіх вагових коефіцієнтів НМ, $\mathbf{v}$ – вектор, компонентами якого є уявні частини всіх вагових коефіцієнтів НМ

3. Метод зворотного поширення похибки навчання ШКНМ

У процесі навчання на кожній ітерації будемо змінювати ваговий вектор у напрямку антиградієнта *E*:

$$\mathbf{w}^{r+1} = \mathbf{w}^r + \Delta\, \mathbf{w}^r , \tag{2}$$

де *r* – номер ітерації і $\Delta\mathbf{w}^r = -\eta_r \operatorname{grad} E(\mathbf{u}^r, \mathbf{v}^r)$.

Нехай $s_{kl}^t = \sum_j w_{jkl}^t z_{jkl}^t + w_{0kl}^t$, $a_{kl}^t = \operatorname{Re} s_{kl}^t$, $b_{kl}^t = \operatorname{Im} s_{kl}^t$, $f(z) = g(x,y) + i\, h(x,y)$. Запишемо компоненти градієнта по вагам вихідного шару:

$$\frac{\partial E}{\partial u_{jkl}} = \sum_t \left(\frac{\partial E}{\partial g_{kl}^t} \left(\frac{\partial g_{kl}^t}{\partial a_{kl}^t}\frac{\partial a_{kl}^t}{\partial u_{jkl}} + \frac{\partial g_{kl}^t}{\partial b_{kl}^t}\frac{\partial b_{kl}^t}{\partial u_{jkl}} \right) + \frac{\partial E}{\partial h_{kl}^t} \left(\frac{\partial h_{kl}^t}{\partial a_{kl}^t}\frac{\partial a_{kl}^t}{\partial u_{jkl}} + \frac{\partial h_{kl}^t}{\partial b_{kl}^t}\frac{\partial b_{kl}^t}{\partial u_{jkl}} \right) \right), \tag{3}$$

$$\frac{\partial E}{\partial v_{jkl}}=\sum_t\left(\frac{\partial E}{\partial g_{kl}^t}\left(\frac{\partial g_{kl}^t}{\partial a_{kl}^t}\frac{\partial a_{kl}^t}{\partial v_{jkl}}+\frac{\partial g_{kl}^t}{\partial b_{kl}^t}\frac{\partial b_{kl}^t}{\partial v_{jkl}}\right)+\frac{\partial E}{\partial h_{kl}^t}\left(\frac{\partial h_{kl}^t}{\partial a_{kl}^t}\frac{\partial a_{kl}^t}{\partial v_{jkl}}+\frac{\partial h_{kl}^t}{\partial b_{kl}^t}\frac{\partial b_{kl}^t}{\partial v_{jkl}}\right)\right). \tag{4}$$

Наведемо розрахункові формули для частинних похідних у (3)-(4) (для спрощення запису не будемо вказувати у наступних формулах індекс t):

$$\frac{\partial \mathrm{E}}{\partial g_{kl}}=g_{kl}-\mathrm{Re}\,d_k,\ \frac{\partial \mathrm{E}}{\partial h_{kl}}=h_{kl}-\mathrm{Im}\,d_k, \tag{5}$$

$$\frac{\partial a_{kl}}{\partial u_{jkl}}=x_{jkl},\ \frac{\partial a_{kl}}{\partial u_{0kl}}=1,\ \frac{\partial b_{kl}}{\partial u_{jkl}}=y_{jkl},\ \frac{\partial a_{kl}}{\partial u_{0kl}}=0, \tag{6}$$

$$\frac{\partial a_{kl}}{\partial v_{jkl}}=-y_{jkl},\ \frac{\partial a_{kl}}{\partial v_{0kl}}=0,\ \frac{\partial b_{kl}}{\partial v_{jkl}}=x_{jkl},\ \frac{\partial a_{kl}}{\partial v_{0kl}}=1. \tag{7}$$

Перейдемо тепер до вибору функції активації. Для дійсних ШНМ частіше за все застосовується логістична сигмоїда $f(x)=\dfrac{1}{1+e^{-x}}$ або тангенс гіперболічний $\mathrm{th}\,x=\dfrac{e^x-e^{-x}}{e^x+e^{-x}}$ (іноді з деякими додатковими параметрами). На жаль, якщо розглядати ці функції як функції комплексної змінної, то вони є розривними, а тому не можуть бути використані для навчання комплексних НМ. Цих недоліків позбавлена раціональна сигмоїда $f(z)=\dfrac{z}{|z|+1}$. Для неї маємо $f(z)=\ g(x,y)+$ $+\ i\ h(x,y)$, де $g(x,y)=\dfrac{x}{\sqrt{x^2+y^2}+1}$, $h(x,y)=\dfrac{y}{\sqrt{x^2+y^2}+1}$. Слід зауважити, що значення, які приймає раціональна сигмоїда, лежать всередині одиничного кола з центром у початку координат. Крім того, раціональна сигмоїда пропорційно стискає дійсну і уявну частини вхідного аргументу, і має властивість підсилювати „слабкі” вхідні сигнали і послаблювати „сильні”. Використовуючи раціональну сигмоїду, отримаємо наступні формули для частинних похідних:

$$\frac{\partial g_{kl}}{\partial a_{kl}}=\frac{b_{kl}^2+|s_{kl}|}{|s_{kl}|\left(|s_{kl}|+1\right)^2},\ \frac{\partial h_{kl}}{\partial b_{kl}}=\frac{a_{kl}^2+|s_{kl}|}{|s_{kl}|\left(|s_{kl}|+1\right)^2},\ \frac{\partial h_{kl}}{\partial a_{kl}}=\frac{\partial g_{kl}}{\partial b_{kl}}=-\frac{a_{kl}b_{kl}}{|s_{kl}|\left(|s_{kl}|+1\right)^2}. \tag{8}$$

Значення похідних $\dfrac{\partial E}{\partial u_{jkl}}$ і $\dfrac{\partial E}{\partial v_{jkl}}$, розраховані по формулам (3)-(8) дозволяють обчислити корекції Δu_{jkl} і Δv_{jkl} для нейронів останнього (зовнішнього) шару. Покажемо, як за їх допомогою можна обчислити корекції вагових коефіцієнтів інших шарів КШНМ. Для останнього шару маємо:

$$\frac{\partial E}{\partial x_{jkl}} = \sum_t \left(\frac{\partial E}{\partial g_{kl}^t} \left(\frac{\partial g_{kl}^t}{\partial a_{kl}^t} \frac{\partial a_{kl}^t}{\partial x_{jkl}} + \frac{\partial g_{kl}^t}{\partial b_{kl}^t} \frac{\partial b_{kl}^t}{\partial x_{jkl}} \right) + \frac{\partial E}{\partial h_{kl}^t} \left(\frac{\partial h_{kl}^t}{\partial a_{kl}^t} \frac{\partial a_{kl}^t}{\partial x_{jkl}} + \frac{\partial h_{kl}^t}{\partial b_{kl}^t} \frac{\partial b_{kl}^t}{\partial x_{jkl}} \right) \right)$$

$$\frac{\partial E}{\partial y_{jkl}} = \sum_t \left(\frac{\partial E}{\partial g_{kl}^t} \left(\frac{\partial g_{kl}^t}{\partial a_{kl}^t} \frac{\partial a_{kl}^t}{\partial y_{jkl}} + \frac{\partial g_{kl}^t}{\partial b_{kl}^t} \frac{\partial b_{kl}^t}{\partial y_{jkl}} \right) + \frac{\partial E}{\partial h_{kl}^t} \left(\frac{\partial h_{kl}^t}{\partial a_{kl}^t} \frac{\partial a_{kl}^t}{\partial v_{jkl}} + \frac{\partial h_{kl}^t}{\partial b_{kl}^t} \frac{\partial b_{kl}^t}{\partial y_{jkl}} \right) \right)$$

В останніх двох формулах частинні похідні $\dfrac{\partial E}{\partial g_{kl}}, \dfrac{\partial E}{\partial h_{kl}}, \dfrac{\partial g_{kl}}{\partial a_{kl}}, \dfrac{\partial g_{kl}}{\partial b_{kl}}, \dfrac{\partial h_{kl}}{\partial a_{kl}}, \dfrac{\partial h_{kl}}{\partial b_{kl}}$ вже розраховані за формулами (6)-(8). Інші частинні похідні мають такий вигляд

$$\frac{\partial a_{kl}}{\partial x_{jkl}} = u_{jkl},\ \frac{\partial b_{kl}}{\partial x_{jkl}} = v_{jkl},\ \frac{\partial a_{kl}}{\partial y_{jkl}} = -v_{jkl},\ \frac{\partial b_{kl}}{\partial y_{jkl}} = u_{jkl}.$$

Але частинні похідні по вхідних значеннях x_{jkl} і y_{jkl} для останнього шару співпадають по змісту з похідними функції похибки E по дійсній і уявній частинам відповідних виходів для попереднього шару. Тому

$$\frac{\partial E}{\partial g_{j,l-1}} = \sum_k \frac{\partial E}{\partial x_{jkl}},\ \frac{\partial E}{\partial h_{j,l-1}} = \sum_k \frac{\partial E}{\partial y_{jkl}}. \tag{9}$$

Формули (9) є аналогом (5) для попередніх шарів і забезпечують перехід від розрахунків координат градієнта поточного шару до розрахунків відповідних координат попереднього шару (метод швидкого обчислення градієнта). Отриманий алгоритм корекції вагових коефіцієнтів за формулами (2)-(9) є комплексною модифікацією алгоритму зворотного поширення похибки, наведеного в [1-2].

При практичному використанні комплексної ШНМ постає питання вибору у (2) множника η_r (коефіцієнта швидкості навчання). Традиційні підходи до знаходження η_r, як розв'язку задачі одновимірної оптимізації є неприйнятними, оскільки вони вимагають численних обчислень похибки мережі *E*. Тому можна покласти $\eta_r = \eta$, де η – деяке наперед вибране число з відрізка $[1/100;1]$. Крім того, для вибору η_r можна запропонувати ті самі підходи, що і у роботі [2].

Навчання КШНМ із функцією помилки вигляду (1) вимагає використання значних об'ємів додаткової пам'яті (одне комплексне число на кожний параметр мережі). Тому для КШНМ із великою кількістю НКНЕ на практиці можна подавати вхідні вектори у випадковому порядку і обмежитися обчисленням градієнта помилки мережі *E* лише по поточному елементу $\left(\mathbf{z}^t, \mathbf{d}^t\right)$ навчальної множини. В цьому випадку формули (3)-(4) набувають простішого вигляду:

$$\frac{\partial E}{\partial u_{jkl}} = \frac{\partial E}{\partial g_{kl}}\left(\frac{\partial g_{kl}}{\partial a_{kl}}\frac{\partial a_{kl}}{\partial u_{jkl}} + \frac{\partial g_{kl}}{\partial b_{kl}}\frac{\partial b_{kl}}{\partial u_{jkl}}\right) + \frac{\partial E}{\partial h_{kl}}\left(\frac{\partial h_{kl}}{\partial a_{kl}}\frac{\partial a_{kl}}{\partial u_{jkl}} + \frac{\partial h_{kl}}{\partial b_{kl}}\frac{\partial b_{kl}}{\partial u_{jkl}}\right), \tag{10}$$

$$\frac{\partial E}{\partial v_{jkl}} = \frac{\partial E}{\partial g_{kl}}\left(\frac{\partial g_{kl}}{\partial a_{kl}}\frac{\partial a_{kl}}{\partial v_{jkl}} + \frac{\partial g_{kl}}{\partial b_{kl}}\frac{\partial b_{kl}}{\partial v_{jkl}}\right) + \frac{\partial E}{\partial h_{kl}}\left(\frac{\partial h_{kl}}{\partial a_{kl}}\frac{\partial a_{kl}}{\partial v_{jkl}} + \frac{\partial h_{kl}}{\partial b_{kl}}\frac{\partial b_{kl}}{\partial v_{jkl}}\right) \tag{11}$$

і розрахунки проводяться за формулами (2), (10)-(11), (5)-(9).

Слід зазначити, для навчання комплексних ШНМ можна використовувати різноманітні модифікації алгоритму зворотного поширення, подібні до тих, які наведені в [1-2] для дійсних ШНМ.

4. Висновки

Комп'ютерне моделювання свідчить, що розглянутий у роботі метод навчання КШНМ з успіхом може бути використаний для апроксимації функцій дійсної та комплексної змінної. Подальшого дослідження потребують питання вибору ефективних функцій активації КНЕ та коефіцієнту швидкості навчання η.

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. Бодянський. – Харків: ТОВ «Компанія СМІТ», 2006. – 404 с.
3. Айзенберг, Н. Н. Многозначная пороговая логика / Н. Н. Айзенберг, Ю. Л. Иваськив. – К. : Наук. думка, 1977. – 145 с.