

УДК 004.42:004.032.26

Л.І. КОРОТКА

Державний вищий навчальний заклад "Український державний хіміко-технологічний університет"

ФУНКЦІОНАЛЬНА ПІДСИСТЕМА РАЦІОНАЛЬНОГО ВИБОРУ АРХІТЕКТУРИ НЕЙРОННОЇ МЕРЕЖІ

У роботі розроблена інформаційна підсистема, яка дозволяє обрати архітектуру слоїстої нейронної мережі та визначитися з її параметрами: кількістю шарів, функцією активації, кількістю епох, похибкою мережі, кроком навчання. Функції інформаційної системи дозволяють отримувати навчальну вибірку для нелінійних систем, вводити за необхідності нестандартну функцію активації та виконувати її чисельне диференціювання для подальшого використання у алгоритмі зворотного поширення помилки. Передбачена можливість динамічного додавання нейронів у шарі.

Ключові слова: штучна нейронна мережа, архітектура нейронної мережі, алгоритм зворотного розповсюдження помилки.

Л.И. КОРОТКАЯ

Государственное высшее учебное заведение "Украинский государственный химико-технологический университет"

ФУНКЦИОНАЛЬНАЯ ПОДСИСТЕМА РАЦИОНАЛЬНОГО ВЫБОРА АРХИТЕКТУРЫ НЕЙРОННОЙ СЕТИ

В работе разработана информационная подсистема, которая позволяет выбрать архитектуру слоистой нейронной сети и определиться с её параметрами: количеством слоёв, функцией активации, количеством эпох, ошибкой сети, шагом обучения. Функции информационной системы позволяют получать обучающую выборку для нелинейных систем, вводит при необходимости нестандартную функцию активации и выполнять её численное дифференцирование для дальнейшего использования в алгоритме обратного распространения ошибки. Предусмотрена возможность динамического добавления нейронов в слое.

Ключевые слова: искусственная нейронная сеть, архитектура нейронной сети, алгоритм обратного распространения ошибки.

L.I. KOROTKA

State Higher Educational Institution "Ukrainian State University of Chemical Technology"

FUNCTIONAL SUBSYSTEM OF RATIONAL CHOICE ARCHITECTURES OF THE NEURAL NETWORK

The information subsystem is developed in the work, which allows choosing the architecture of the layered neural network and determining its parameters: the number of layers, the activation function, the number of epochs, the network error, the learning step. The functions of the information system make it possible to obtain a learning sample for nonlinear systems, to introduce, if necessary, a nonstandard activation function and to perform its numerical differentiation for further use in the algorithm for back propagation of the error. It is possible to dynamically add neurons to the layer.

Keywords: artificial neural network, neural network architecture, back error propagation algorithm.

Постановка проблеми

Сучасний стан розвитку інформаційних систем (ІС) характеризується достатньо широким використанням Internet/Intranet-технологій у синтезі з елементами штучного інтелекту. Технології обчислювального інтелекту застосовуються у різних предметних областях, у тому числі штучні нейронні мережі (НМ), як елементи цієї технології. Використання НМ має достатні переваги, але пов'язане з деякими труднощами, наприклад, навчанням мереж та вибором їх параметрів.

Аналіз останніх досліджень і публікацій

Автор не претендує на вичерпний аналіз існуючого програмного забезпечення, яке дозволяє проектувати, навчати та використовувати нейронні мережі для вирішення різних прикладних задач. Звернемо увагу тільки на те, що нейромережеві середовища, як правило, не працюють одночасно з декількома НМ та не використовують нестандартні функції активації.

Мета дослідження

Об'єктом дослідження є моделювання процесу навчання обчислювальних структур – штучних нейронних багат шарових мереж прямого розповсюдження при розв'язанні систем нелінійних рівнянь. Метою роботи є створення web-орієнтованої функціональної підсистеми визначення раціональної архітектури нейронних мереж для розв'язання прикладних задач. Постановка задачі: необхідно провести навчання декількох штучних нейронних мереж алгоритмом зворотного розповсюдження помилки (ЗРП) з неперервними функціями активації (ФА) з метою вибору однієї з НМ для подальшого використання.

Математична постановка: нехай задано навчальну множину $\{\tilde{x}_i, \tilde{y}_i\}, i = \overline{1, k}$. Тут $\tilde{x}_i \in R^n$ та $\tilde{y}_i \in R^l$ – тут відповідно вхідний вектор та вектор вказівок вчителя i -ого прикладу. Нейромережева модель задається функціоналом:

$$J(w) = \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^l [y_{ij}(w) - \tilde{y}_{ij}]^2. \quad (1)$$

Для навчання нейронної мережі застосовується алгоритм зворотного поширення помилки, який для налаштування вагових коефіцієнтів використовує метод градієнтного спуску:

$$w^* = -\eta \cdot \frac{\partial J}{\partial w}. \quad (2)$$

У процесі навчання необхідно визначити вектор ваг w^* такий, щоб функціонал (1) приймав мінімальне значення:

$$w^* = \arg \min_{w \in R^e} J(w), \quad (3)$$

де $y_{ij}(w)$ – зразок, який виданий мережею; w – матриця синаптичних ваг; η – коефіцієнт швидкості навчання НМ, значення якого береться з інтервалу $0 < \eta < 1$; \tilde{y}_{ij} – еталонне значення.

Викладення основного матеріалу дослідження

Для досягнення поставленої мети та виходячи з постановки задачі, функції системи, які вона повинна виконувати, можна сформулювати наступним чином:

1. Отримати з файлу навчальну вибірку $\{\tilde{x}_i, \tilde{y}_i\}, i = \overline{1, k}$ та зарезервувати третину тестових зразків для нелінійної системи рівнянь.
2. Задати для подальшого визначення раціональної архітектури нейронної мережі для кожної з них параметри навчання, а саме: вид функції активації, кількість шарів, похибку мережі, коефіцієнт навчання, кількість вхідних та вихідних елементів. Зауважимо, що система одночасно працює з чотирма нейронними мережами.
3. Провести одночасне навчання НМ алгоритмом зворотного розповсюдження помилки.
4. Протестувати навчені нейронні мережі.
5. Задати (за необхідності) нестандартні ФА та виконати їх чисельне диференціювання для подальшого використання у алгоритмі ЗРП.
6. Додавати у процесі навчання нейрони у прихованому шарі НМ.
7. Отримати за необхідності навчальну вибірку для нелінійної системи рівнянь, причому передбачається можливість завдання системи безпосереднім введенням рівнянь з клавіатури.

Як відомо з теорії нейронних мереж, вибір топології мережі визначається розв'язуваною задачею. Для апроксимації функцій багатьох змінних достатньо добре зарекомендували себе слоїсті нейронні мережі [1]. Підкреслимо, щодо топології розглянутих у роботі штучних нейронних мереж, то було проведено процес навчання для слоїстих НМ. У якості алгоритму навчання використано алгоритм зворотного поширення помилки.

Для досягнення поставленої мети необхідно звернути увагу на деякі проблемні аспекти, з якими стикається кожний проектувальник нейронної мережі. Розглянемо їх детальніше. Від якості навчальної вибірки залежить якість навчання нейронної мережі. Дуже суттєвим є об'єм вибірки та її зразки, які повинні якомога краще описувати середовище. Тому у роботі відслідковуються ситуації достатньої або недостатньої кількості навчальних зразків [2–3]. Як відомо при малому об'ємі вибірки нейронна мережа, як правило, буде недостатньо навчена, а при дуже великій кількості навчальних зразків може відбутися так зване "перенавчання", тому доцільно робити перевірку об'єму вибірки, наприклад, за формулою:

$$N \geq \frac{m_0 \cdot m_1}{\varepsilon_{нс}}, \quad (4)$$

де m_0, m_1 – відповідно розміри вхідного та вихідного шарів; $\varepsilon_{нс}$ – похибка мережі; N – об'єм навчальної вибірки.

Для попередження ситуацій перенавчання нейронної мережі у роботі використано механізм контрольної крос-перевірки [3, 5].

Оцінювання кількості нейронів L_w у прихованих шарах однорідних нейронних мереж визначається за формулою [4]:

$$\frac{m \cdot N}{1 + \log_2 N} \leq L_w \leq m \left(\frac{N}{m} + 1 \right) \cdot (n + m + 1) + m,$$

де n – розмірність вхідного сигналу; m – розмірність вихідного сигналу.

Зауважимо, що згідно теореми про повноту [4], у роботі розглядаються тільки одно- та двошарові нейронні мережі. Що стосується таких параметрів навчання, як коефіцієнт навчання та похибка нейронної мережі, то вони визначаються проектувальником нейронної мережі. Значення коефіцієнта можна варіювати у відомому діапазоні. Очевидно, що невелике його значення сповільнює процес навчання, але й велике значення може призвести до того, що можливий локальний мінімум буде пропущено. Похибка нейронної мережі безпосередньо залежить від об'єму вибірки згідно формули (3). Достатньо детально евристичні моменти навчання нейронної мережі алгоритмом зворотного поширення помилки описані в [5].

Останнім аспектом, який буде розглянуто і який представляється автору важливим при роботі з нейронними мережами, є вибір ФА. Очевидно, що можна вибирати як стандартні функції, такі як логістичні (сигмоїд або гіперболічний тангенс), так і нестандартні (наприклад, у роботі використовується експонента та частина "хвилі" синусоїди). Крім того передбачено введення з консолі користувачем нестандартної ФА, але слід зауважити, що згідно теореми про повноту, функція активації повинна бути диференційованою. У роботі використовується чисельне диференціювання таких ФА для подальшого використання у алгоритмі ЗРП.

Після описаних аспектів, з якими обов'язково стикаються проектувальники НМ та необхідного математичного апарату, перейдемо до опису створюваної ІС.

Для функціональної підсистеми було використано наступне програмне забезпечення:

- скриптова мова програмування для написання web-сторінок – JavaScript;
- бібліотека JavaScript з відкритим сирцевим кодом – jQuery;
- скриптова мова програмування для генерації HTML-сторінок – php;
- відкритий web-сервер Інтернет – Apache HTTP-сервер;
- мова для структурування та подання вмісту – HTML;
- спеціальна мова відображення web-сторінок, написаних мовами розмітки даних – CSS3;
- вільний набір інструментів для створення сайтів та web-додатків – Twitter Bootstrap.

Перед безпосереднім використанням НМ необхідно навчити мережу узагальнювати отримані нею знання, але цей процес не є простим. Від проектувальника мережі, як зазначалося раніше, залежить багато факторів, у тому числі вибір архітектури НМ та її параметрів. Тому у роботі є можливість працювати з декількома нейронними мережами та обирати різні ФА (а саме чотири). Звісно, що є можливість варіювання параметрів навчання.

Для наочності процесу навчання НМ алгоритмом ЗРП будуються графіки значень похибки. Графіки дозволяють особі, яка приймає рішення, спостерігати процес навчання цих мереж. Саме графіки дають можливість бачити якість навчання мережі, у тому числі і її перетренування.

Результати чисельних експериментів. Як зазначалося раніше, у роботі використано чотири ФА: логістичний сигмоїд, гіперболічний тангенс, "хвилю" синусоїди на заданому інтервалі та експоненційну функцію. Було проведено навчання нейронних мереж, результати якого наведено у табл. 1. У роботах показано, що [1, 3] сигмоїдальні функції активації дають практично однакові результати, а якщо порівнювати їх ФА s -подібного типу, то і результати будуть відрізнятися. Слід зазначити, що вибір тієї чи іншої мережі залишається за особою, яка приймає рішення.

Таблиця 1

Результати навчання нейронних мереж алгоритмом ЗРП

Архітектура НМ	Функція активації	Кількість епох	Похибка НМ
4-4-3	Логістичний сигмоїд	534	0,00043
4-6-3	Частина «хвилі» синусоїди на заданому інтервалі	176	0,00005
4-4-4-3	Гіперболічний тангенс	341	0,00046

За результатами дослідження чисельних експериментів навчання нейронних мереж можна стверджувати, що розроблена підсистема дозволяє:

- створювати вибірку для нейронних мереж будь-якої складності та об'єму;

- використовувати чисельне диференціювання неперервних функцій активації;
- створювати та налаштовувати штучні нейронні мережі різних архітектур, з різними параметрами навчання;
- здійснювати візуалізацію процесу навчання нейронних мереж за допомогою графіків змінення похибки навчання;
- додавати динамічно нейрони у приховані шари;
- визначати особі, яка приймає рішення, раціональну архітектуру НМ для розв’язання прикладних задач.

Слід зупинитись на головних етапах навчання нейронних мереж. В тому випадку коли задані всі параметри для навчання мережі, нормалізовано вхідні дані та задано похибку навчання, можна безпосередньо перейти до процесу навчання мережі. Реалізовані функції ІС дозволяють будувати чотири графіки, на яких відображається динамічний процес навчання НМ та змінення значень похибки мережі у ході її навчання (рис. 1).

Після етапу навчання можна переходити до тестування мережі. Цей етап дозволяє визначити наскільки НМ здатна до узагальнення. Функціональна підсистема дозволяє спостерігати цей етап, будувати графіки тестування НМ аналогічні, що представлені на рис. 1 та виводити додаткову інформацію.

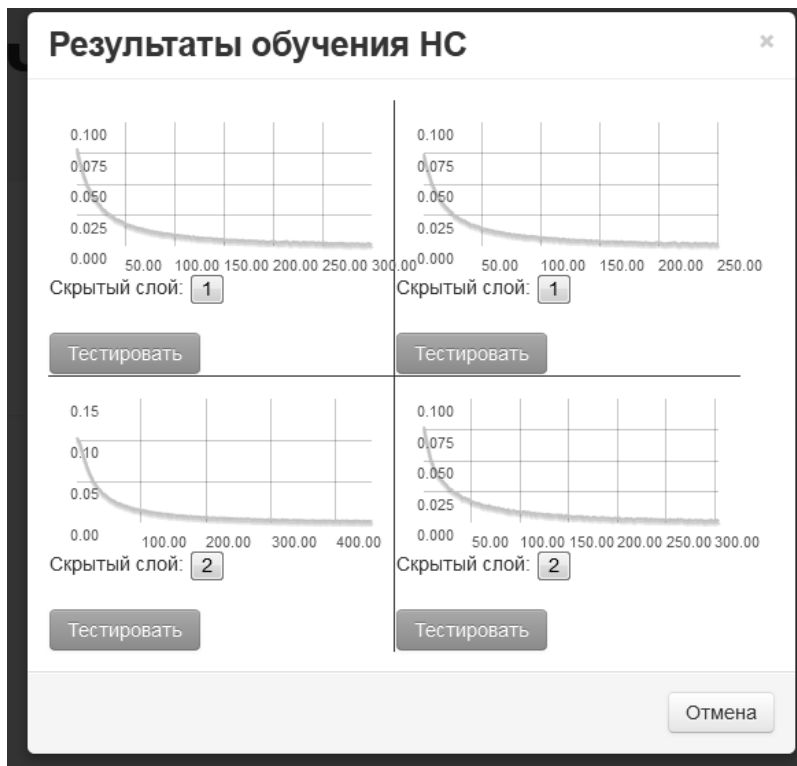


Рис. 1. Вікно вибору навченої НМ для тестування

Таким чином, головні етапи навчання нейронних мереж реалізовано у web-орієнтованій підсистемі. Функції системи дозволяють одночасно працювати з кількома нейронними мережами, але вибір та переваги щодо конкретної НМ залишається за спеціалістом у даній предметній області. Тільки він може визначити наскільки та чи інша мережа здатна до якісного узагальнення.

Як відомо, навчена нейронна мережа володіє пам’яттю, заради якої і потрібен такий довгий та не простий шлях навчання. У роботі передбачено збереження синаптичних вагових коефіцієнтів у окремий файл для подальшого використання.

Нейронні мережі, як елемент технології обчислювального інтелекту, мають свої переваги та недоліки, але вони дуже добре себе зарекомендували як універсальні апроксимуючі системи.

Висновки

Підсумовуючи вищевикладений матеріал можна відмітити, що розроблений нейроемулятор дозволяє: створювати, проектувати та налаштовувати штучні нейронні мережі різної архітектури; за необхідності отримувати вибірку для нейронних мереж будь-якої складності та розміру; використовувати чисельне диференціювання неперервних функцій активації; здійснювати візуалізацію процесу навчання та тестування нейронних мереж; визначатися особі, що приймає рішення, щодо раціональної архітектури мережі.

Список використаної літератури

1. Зеленцов Д.Г. Нейронные сети как средство модификации метода скользящего допуска / Д.Г. Зеленцов , Л.И. Короткая // Восточно-европейский журнал передовых технологий. – 2011. – № 4/4 (52). – С. 21–24.
2. Каллан Р. Основные концепции нейронных сетей / Р. Каллан. – М.: Издательский дом "Вильямс", 2001. – 288 с.
3. Короткая Л.И. Использование нейронных сетей при численном решении некоторых систем дифференциальных уравнений / Л.И. Короткая // Восточно-европейский журнал передовых технологий. – 2011. – № 3/4 (51). – С. 24–27.
4. Круглов В.В. Нечеткая логика и искусственные нейронные сети / В.В. Круглов, М.И. Дли, Р.Ю. Голунов. – М.: Физматлит, 2001. – 221 с.
5. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – М.: Издательский дом "Вильямс", 2006. – 1104 с.