

ПОБУДОВА ОПТИМІЗОВАНОЇ БАГАТОШАРОВОЇ НЕЙРОННОЇ МЕРЕЖІ В МЕЖАХ НЕЛІНІЙНОЇ МОДЕЛІ УЗАГАЛЬНЕНОЇ ПОХИБКИ

Роман Пелещак¹, Василь Литвин², Іван Пелещак³,
Вікторія Висоцька⁴, Оксана Черняк⁵

¹⁻⁴ Національний університет “Львівська політехніка”

⁵ Дрогобицький державний педагогічний університет імені Івана Франка

¹ rpeleshchak@ukr.net, ORCID: 0000-0002-0536-3252

² vasyly.v.lytvyn@lpnu.ua, ORCID: 0000-0002-9676-0180

³ peleshchakivan@gmail.com, ORCID: 0000-0002-7481-8628

⁴ victoria.a.vysotska@lpnu.ua, ORCID: 0000-0001-6417-3689

⁵ oksana2895@gmail.com, ORCID: 0000-0001-6034-1016

© Пелещак Р., Литвин В., Пелещак І., Висоцька В., Черняк О., 2021

У роботі запропоновано спосіб оптимізації структури багатошарової нейронної мережі на основі мінімізації нелінійної узагальненої похибки, яка ґрунтується на принципі мінімальної довжини опису. Відповідно до цього принципу, узагальнена похибка мережі під час роботи з новими даними визначається похибкою апроксимації даних нейронною мережею у нелінійному наближенні та похибкою опису моделі. З умови мінімуму узагальненої похибки мережі виведено вирази для обчислення оптимального розміру мережі (кількість синаптичних зв'язків та кількість нейронів у прихованих шарах). Побудовано графічні залежності узагальненої похибки мережі від кількості синаптичних зв'язків між нейронами за різних значень вхідних образів і фіксованої кількості навчальних прикладів та графічні залежності оптимальної кількості синаптичних зв'язків від кількості навчальних прикладів за різних значень вхідних образів. На основі співвідношень для оптимальної кількості синаптичних зв'язків між нейронами та оптимальної кількості нейронів у прихованих шарах оцінено ступінь складності навчання нейронної мережі.

Ключові слова: нейронна мережа; узагальнена похибка; оптимізація розміру.

Вступ. Постановка проблеми

Сьогодні ведуться інтенсивні дослідження щодо застосування нейронних мереж для розв'язання широкого класу задач Data Mining (ідентифікація нестационарних хаотичних процесів, кластеризація, класифікація, інтелектуальне керування, діагностика станів біосистем, прогнозування, емуляція та розпізнавання вхідних образів). Для розв'язування вказаних задач із високою точністю необхідно знати оптимальну архітектуру нейронних мереж, тобто оптимальну кількість синаптичних зв'язків між нейронами в різних шарах та оптимальну кількість нейронів [1, 2].

У випадку моделювання нейронної мережі кількість її параметрів, як правило, велика, зокрема, розмір мережі ($N_{\lambda_{ij}}$ – загальна кількість синаптичних зв'язків, а індекс λ_{ij} описує значення

синаптичних зв'язків між i -м та j -м нейронами; n_k – кількість нейронів у прихованих шарах) співвідноситься із розміром навчальної вибірки N_p ($N_{\lambda_{ij}} \lesssim N_p$), тобто кількість параметрів залежить від кількості навчальних даних і кількості вхідних образів (N_{in}).

Аналіз останніх досліджень та публікацій

Одним із головних критеріїв ефективності функціонування багатошарової нейронної мережі є узагальнена похибка ε [3–5]. За допомогою цієї похибки можна оцінити точність нейронної мережі, що здатна класифікувати дані, не представлені у навчальній вибірці. Оптимізацію узагальненої похибки, яка складається із суми похибки апроксимації ε_a (емпірична похибка) та похибки, пов'язаної зі складністю моделі ε_m (морфологією моделі нейронної мережі), визначає принцип мінімальної довжини опису [3].

$$\min_{N_{\lambda_{ij}}} \left\{ -\ln P(N_p | (N_{\lambda_{ij}}; \lambda_{ij})) - \ln P(N_{\lambda_{ij}}; \lambda_{ij}) \right\} = \min_{N_{\lambda_{ij}}} \{ \varepsilon_a + \varepsilon_m \}, \quad (1)$$

де перший доданок відображає похибку апроксимації (умовна ймовірність). Що менша похибка апроксимації, то менше бітів інформації необхідно для виправлення передбачень моделі. Зокрема, в задачі апроксимації функції припускають, що множина взаємопов'язаних даних $\{\bar{x}_\alpha, \bar{y}_\alpha\}$ описується деякою невідомою функцією, яку і потрібно встановити, але їх істинні значення спотворені випадковим гауссівським шумом. Отже, умовна ймовірність множини даних $\{\bar{x}_\alpha, \bar{y}_\alpha\}$ для моделі нейронної мережі $\left\{ \bar{y}(\bar{x}_\alpha; N_{\lambda_{ij}}; n_k; \lambda_{ij}) \right\}$, яка залежить від налаштовуваних параметрів $N_{\lambda_{ij}}$, n_k , λ_{ij} , має гауссівський розподіл

$$P(y_\alpha | (N_{\lambda_{ij}}; \lambda_{ij})) \propto \exp \left[\frac{-(y_\alpha - y(\bar{x}_\alpha; N_{\lambda_{ij}}; n_k; \lambda_{ij}))^2}{2\sigma^2} \right], \quad (2)$$

де σ^2 – дисперсія. Необхідно зазначити, що правильна апроксимація зашумленої функції $\left\{ \bar{y}(\bar{x}_\alpha; N_{\lambda_{ij}}; n_k; \lambda_{ij}) \right\}$ повинна давати похибку, близьку до дисперсії шуму. Тоді із урахуванням (2)

похибка апроксимації $P(N_p | (N_{\lambda_{ij}}; \lambda_{ij}))$ дорівнює

$$P(N_p | (N_{\lambda_{ij}}; \lambda_{ij})) = \sum_{\alpha=1}^{N_p} P(y_\alpha | (N_{\lambda_{ij}}; \lambda_{ij})). \quad (3)$$

Згідно із принципом мінімальної довжини опису (1) можна мінімізувати загальну довжину опису даних на основі апроксимації функції нейронною мережею та опису самої моделі. Якщо модель передбачає всі дані точно, то довжина опису похибки дорівнює нулю. Другий доданок у формулі (1) визначає зміст кількості інформації, необхідної для вибору конкретної моделі з множини із апріорним розподілом ймовірностей $P(N_{\lambda_{ij}}; \lambda_{ij})$.

Зокрема, в роботах [3, 6] похибки апроксимації та опису моделі подано у лінійних наближеннях $\varepsilon_a \sim o\left(\frac{N_{in}}{N_{\lambda_{ij}}}\right)$ та $\varepsilon_m \sim o\left(\frac{N_{\lambda_{ij}}}{N_p}\right)$, де N_{in} – кількість вхідних образів. У такому наближенні формула для узагальненої похибки ε нейронної мережі описувалась виразом

$$\varepsilon \sim \varepsilon_a + \varepsilon_m = \frac{N_{in}}{N_{\lambda_{ij}}} + \frac{N_{\lambda_{ij}}}{N_p}. \quad (4)$$

Необхідно зазначити, що складові ε_a і ε_m по-різному залежать від розміру нейронної мережі (кількості синаптичних зв'язків та нейронів), що припускає можливість вибору оптимального розміру нейронної мережі. Зауважимо, що модель (4) добре описує нейронну мережу в наближеннях

$$\frac{N_{in}}{N_{\lambda_{ij}}} \ll 1 \text{ та } \frac{N_{\lambda_{ij}}}{N_p} \ll 1.$$

У випадках, коли відношення кількості вхідних образів до кількості синаптичних зв'язків становить $\frac{N_{in}}{N_{\lambda_{ij}}} < 1$, а відношення кількості синаптичних зв'язків до кількості навчальних прикладів

становить $\frac{N_{\lambda_{ij}}}{N_p} \lesssim 1$ (якщо кількість синаптичних зв'язків $N_{\lambda_{ij}}$ не перевищує кількості навчальних

прикладів N_p , модель узагальненої похибки (4) неточно описуватиме похибку ε нейронної мережі).

У разі виконання останніх умов похибку апроксимації ε_a необхідно описувати в наближенні

$$\varepsilon_a \sim o\left(\frac{N_{in}}{N_{\lambda_{ij}}}\right), \text{ а похибку опису моделі } \varepsilon_m \text{ – у нелінійному наближенні } \varepsilon_m \sim o\left(\frac{N_{\lambda_{ij}}^2}{N_p^2}\right).$$

$$\varepsilon_a \sim o\left(\frac{N_{in}}{N_{\lambda_{ij}}}\right), \quad \varepsilon_m \sim o\left(\frac{N_{\lambda_{ij}}^2}{N_p^2}\right) \quad (5)$$

Іншою проблемою оптимізації багатошарової нейронної мережі є перенавчання. Нехай навчальна вибірка описується певною функцією, яку нам потрібно відтворити. Під час навчання таку функцію називають вчителем. За обмеженої кількості навчальних прикладів N_p можна побудувати нейронну мережу із мінімальною похибкою навчання, тобто похибкою, яка формується на множині навчальних прикладів. Для цього потрібно взяти нейронну мережу з кількістю синаптичних зв'язків ($N_{\lambda_{ij}}$, де λ_{ij} – синаптичні зв'язки між i -м та j -м нейронами), більшою за кількість навчальних прикладів N_p , тобто ($N_{\lambda_{ij}} > N_p$). Щоб відтворити кожен приклад, маємо N_p рівнянь для $N_{\lambda_{ij}}$ невідомих. Відомо [7], що коли рівнянь N_p менше, ніж невідомих $N_{\lambda_{ij}}$, така система є недовизначеною і допускає нескінченну множину розв'язків. Головна проблема у тому, що не вистачає інформації, щоб вибрати єдине правильне рішення – функцію-вчителя. Вибрана випадково функція спричиняє неточності на нових прикладах, які відсутні в навчальній вибірці, хоч навчальну вибірку відтворює без похибок. Замість того, щоб узагальнити навчальну вибірку, мережа запам'ятала її. Цей ефект називається перенавчанням [3].

Формулювання цілі статті

Зважаючи на зазначене вище, завданням навчання нейронної мережі є не мінімізація похибки навчання, а мінімізація похибки узагальнення ε , визначеної для всіх можливих прикладів у майбутньому. Мережа з мінімальною похибкою узагальнення ε_{\min} максимально точно розпізнаватиме та класифікуватиме образи.

За аналогією із біологічною нейронною мережею похибка апроксимації (навчання) співвідноситься із запам'ятовуванням навчальних даних багатошаровим перцептроном, а похибка опису моделі співвідноситься із узагальненням (стисненням) цих даних. Зазначимо, що запам'ятовування і стиснення навчальних даних відбувається за рахунок зміни вагових коефіцієнтів синаптичних зв'язків та зміни топології нейронів у нейронній мережі.

Завдання мінімізації кількості нейронів і синаптичних зв'язків у нейронних мережах прямого поширення розглянуто у роботі [8] на основі метричних методів розпізнавання.

Метою цієї роботи є оптимізація кількості синаптичних зв'язків та нейронів у прихованих шарах у багатошаровій нейронній мережі в наближенні нелінійної узагальненої похибки ε .

Нелінійна модель оптимальної кількості синаптичних зв'язків та нейронів

Критерієм оптимізації загальної кількості синаптичних зв'язків $N_{\lambda_{ij}}$ вибираємо узагальнену похибку ε багатошарової нейронної мережі. Відповідна математична модель має вигляд

$$\varepsilon(N_{\lambda_{ij}}) \rightarrow \varepsilon_{\min}; N_{\lambda_{ij}} \in [1, \infty). \quad (6)$$

У наближенні нелінійної моделі узагальнену похибку ε з урахуванням умов (5) записуємо у вигляді

$$\varepsilon = \frac{N_{in}}{N_{\lambda_{ij}}} + \frac{N_{\lambda_{ij}}}{N_p} + \frac{N_{\lambda_{ij}}^2}{N_p^2}. \quad (7)$$

Як видно зі співвідношення (7), збільшення кількості синаптичних зв'язків $N_{\lambda_{ij}}$, й відповідно кількості нейронів у прихованих шарах n_k ($k = 1, 2, 3, \dots, m$, де k – номер шару), приводить до зменшення похибки апроксимації навчальних даних, тоді як друга складова ε_m узагальненої похибки ε – до збільшення похибки опису моделі ε_m .

Для визначення мінімального значення узагальненої похибки ε_{\min} , яке відповідає оптимальній кількості синаптичних зв'язків $N_{\lambda_{ij}}^{opt}$, необхідно розв'язати рівняння

$$\frac{\partial \varepsilon}{\partial N_{\lambda_{ij}}} = 0. \quad (8)$$

З урахуванням (7) із рівняння (8) отримаємо вираз для оптимальної кількості синаптичних зв'язків $N_{\lambda_{ij}}^{opt}$

$$N_{\lambda_{ij}}^{opt} = -\frac{N_p}{6} + \frac{N_p^2}{6 \cdot (54N_{in}N_p^2 - N_p^3 + 6\sqrt{3}\sqrt{27N_{in}^2N_p^4 - N_{in}N_p^5})^{\frac{1}{3}}} + \frac{1}{6} \left(54N_{in}N_p^2 - N_p^3 + 6\sqrt{3}\sqrt{27N_{in}^2N_p^4 - N_{in}N_p^5} \right)^{\frac{1}{3}} \quad (9)$$

Як видно з формули (9), оптимальна кількість синаптичних зв'язків $N_{\lambda_{ij}}^{opt}$ нелінійно залежить від кількості вхідних образів N_{in} та кількості навчальних прикладів N_p .

Підставивши $N_{\lambda_{ij}}^{opt}$ (9) у формулу (7), отримаємо мінімальне значення узагальненої похибки $\varepsilon_{\min} = \varepsilon_{|N_{\lambda_{ij}}=N_{\lambda_{ij}}^{opt}}$ багатошарового перцептрона (БШП).

Взаємозв'язок між загальною кількістю синаптичних зв'язків $N_{\lambda_{ij}}$ та кількістю нейронів у прихованих шарах n_k БШП визначається формулою [6]

$$N_{\lambda_{ij}} = N_{in} \cdot n_1 + \sum_{k=1}^{m-1} (n_k \cdot n_{k+1}) + n_m \cdot n_{out}, \quad (10)$$

де n_{out} – кількість нейронів у вихідному шарі.

Маючи вирази (9) та (10) для розрахунку оптимальної кількості синаптичних зв'язків $N_{\lambda_{ij}}^{opt}$, можна визначити оптимальну кількість нейронів n_k^{opt} у прихованих шарах нейромережі.

Для подання довільної функції згідно із теоремою Хехт–Нільсена [9] достатньо тришарової нейромережі (з одним прихованим шаром, $m = 1$). Згідно із формулою (10) з урахуванням співвідношення (9) оптимальна кількість нейронів у одному прихованому шарі n_1^{opt} тришарової нейронної мережі визначатиметься так

$$n_1^{opt} = \frac{-\frac{N_p}{6} + \frac{N_p^2}{6 \cdot (54N_{in}N_p^2 - N_p^3 + 6\sqrt{3}\sqrt{27N_{in}^2N_p^4 - N_{in}N_p^5})^{\frac{1}{3}}} + \frac{1}{6} \left(54N_{in}N_p^2 - N_p^3 + 6\sqrt{3}\sqrt{27N_{in}^2N_p^4 - N_{in}N_p^5} \right)^{\frac{1}{3}}}{N_{in} + n_{out}} \quad (11)$$

Результати числових розрахунків та порівняння із комп'ютерним експериментом

Наведемо результати числових розрахунків у середовищі “Математика” залежності оптимальної кількості синаптичних зв'язків від кількості навчальних прикладів $N_{\lambda_{ij}}^{opt} = f(N_p)$ для чотирьох значень вхідних образів $N_{in} : 4; 10; 20; 50$. Як видно з рис. 1, функціональна залежність $N_{\lambda_{ij}}^{opt} = f(N_p)$ на усьому проміжку $0 \leq N_p \leq 500$ монотонно зростає. Зі збільшенням кількості вхідних образів N_{in} зростає різниця між функціональною залежністю $N_{\lambda_{ij}}^{opt} = f(N_p)$ у лінійній та нелінійній моделях узагальної похибки. Зокрема, якщо $N_p = 500$ і $N_{in} = 50$, у нелінійній моделі оптимальна кількість синаптичних зв'язків на 20 % менша порівняно із лінійною моделлю. Найменшою різниця оптимальної кількості синаптичних зв'язків буде, якщо $N_{in} = 4$.

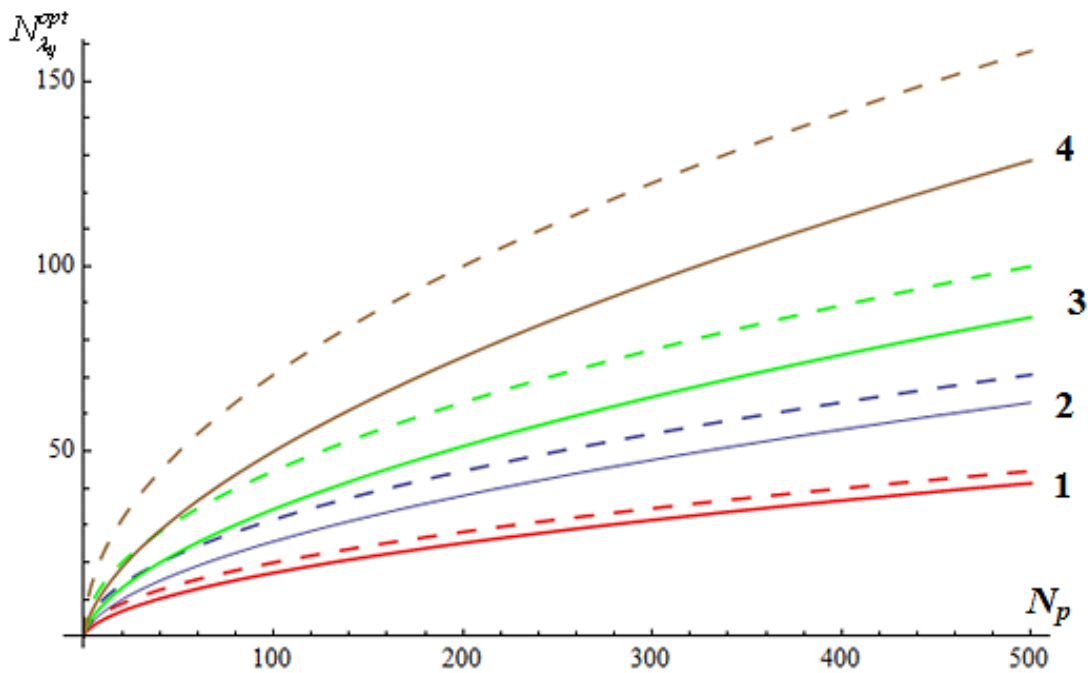


Рис. 1. Залежність оптимальної кількості синаптичних зв'язків від кількості навчальних прикладів.
1 – $N_{in} = 4$; 2 – $N_{in} = 10$; 3 – $N_{in} = 20$; 4 – $N_{in} = 50$; штрихові лінії – лінійне наближення

На рис. 2 подано результати числових розрахунків узагальної похибки залежно від загальної кількості синаптичних зв'язків, тобто $\varepsilon = \varphi(N_{\lambda_{ij}})$ за кількості навчальних прикладів $N_p = 1000$ і $N_{in} : 4; 10; 20; 50$. Як видно з результатів числових розрахунків, такі функціональні залежності $\varepsilon = \varphi(N_{\lambda_{ij}})$ немонотонні. Мінімуму узагальної похибки відповідає оптимальна кількість зв'язків $N_{\lambda_{ij}}^{opt}$, тобто $\varepsilon_{min} = \varphi(N_{\lambda_{ij}}^{opt})$. Зі зростанням кількості вхідних образів N_{in} мінімум узагальної похибки зсувається у бік більшої кількості синаптичних зв'язків. За значень кількості синаптичних зв'язків, які містяться у інтервалі $0 < N_{\lambda_{ij}} < 100$, значення узагальної похибки у нелінійній моделі практично збігається зі значенням у лінійній. Якщо $N_{\lambda_{ij}} \geq 100$, відбувається збільшення узагальної похибки $\varepsilon = \varphi(N_{\lambda_{ij}})$, причому в межах нелінійної моделі швидкість зростання узагальної похибки $\frac{\Delta \varepsilon}{\Delta N_{\lambda_{ij}}}$ більша порівняно зі швидкістю зростання у лінійній.

Порівняння числових розрахунків узагальної похибки в межах нелінійної моделі з комп'ютерним експериментом, виконаним на прикладі оброблення сигналів розподіленою волоконно-

оптичною мережею для розпізнавання динамічних образів з використанням нейронних мереж [10], показало кореляцію між похибкою, отриманою в комп'ютерному експерименті, і одержаними числовими результатами зі зростанням кількості навчальних прикладів (рис. 1, 2).

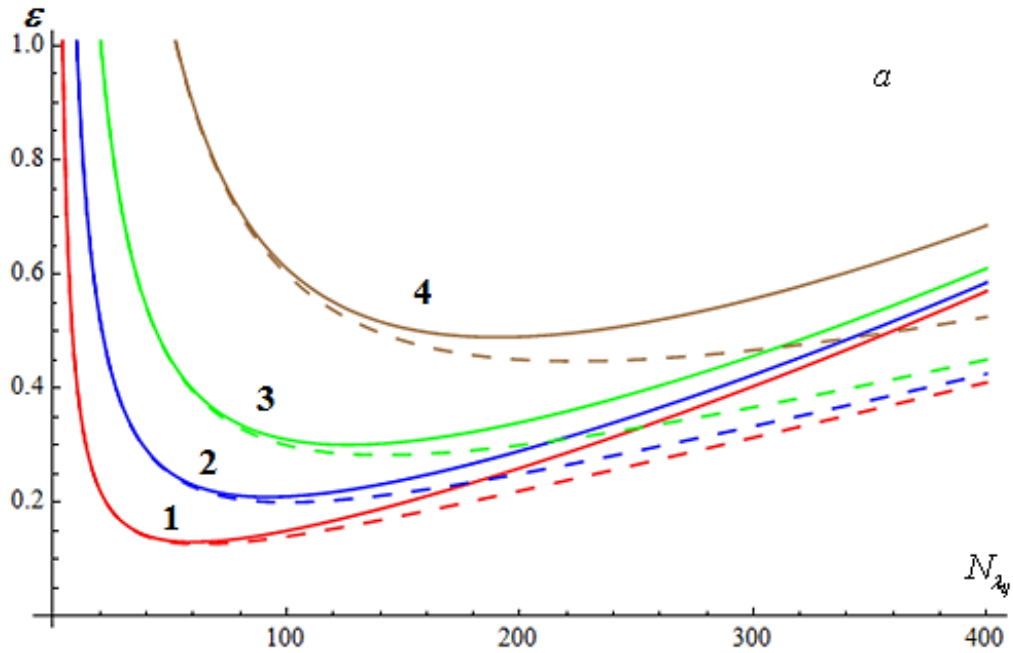


Рис. 2. Залежність узагальненої похибки від кількості синаптичних зв'язків.
1 – $N_{in} = 4$; 2 – $N_{in} = 10$; 3 – $N_{in} = 20$; 4 – $N_{in} = 50$; штрихові лінії – лінійне наближення

На рис. 3 зображено графічні залежності оптимальної кількості нейронів від кількості навчальних прикладів, тобто $n_1^{opt} = \psi(N_p)$, якщо $N_{in} : 4; 10; 20; 50$. Видно, що ці залежності монотонно зростають. Зі збільшенням кількості сенсорних нейронів N_{in} (вхідних образів) зменшується оптимальна кількість нейронів у прихованому шарі тришарової нейронної мережі.

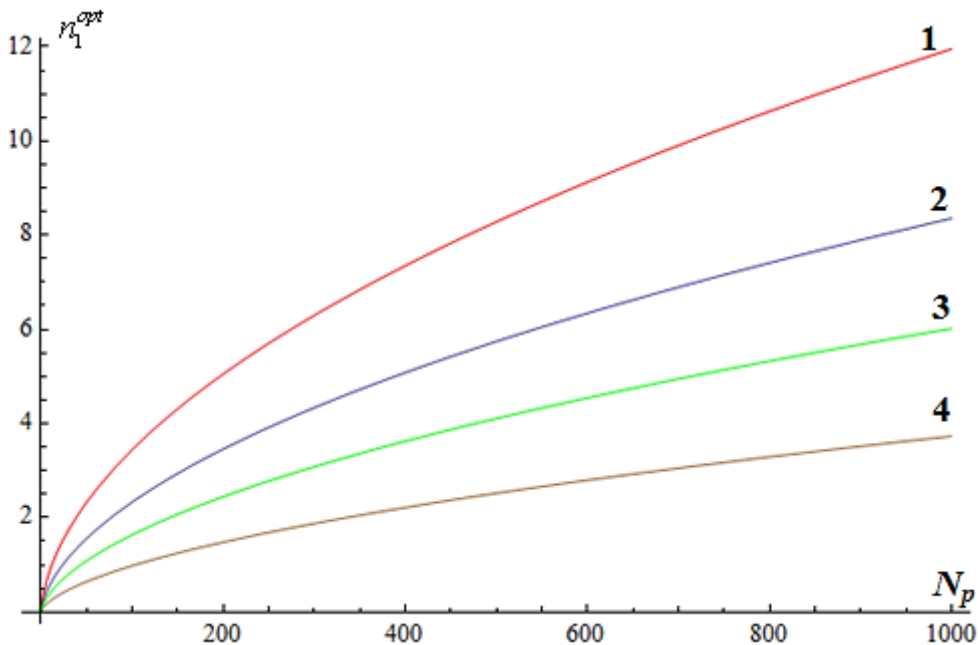


Рис. 3. Залежність оптимальної кількості нейронів у прихованому шарі від кількості навчальних прикладів.
1 – $N_{in} = 4$; 2 – $N_{in} = 10$; 3 – $N_{in} = 20$; 4 – $N_{in} = 50$; штрихові лінії – лінійне наближення

На рис. 4 подано графічну залежність параметра складності навчання (кількість операцій, виконаних нейронною мережею) $C = N_p \times (N_{\lambda_{ij}}^{opt})^2$ від кількості навчальних прикладів N_p , тобто $C = \phi(N_p)$; де $N_{\lambda_{ij}}^{opt}$ визначається з формули (9). Зі збільшенням кількості вхідних образів N_{in} та кількості навчальних прикладів N_p ступінь складності навчання C зростає нелінійно. У лінійній моделі узагальненої похибки зі збільшенням N_p ступінь складності навчання є вищим порівняно із нелінійною.

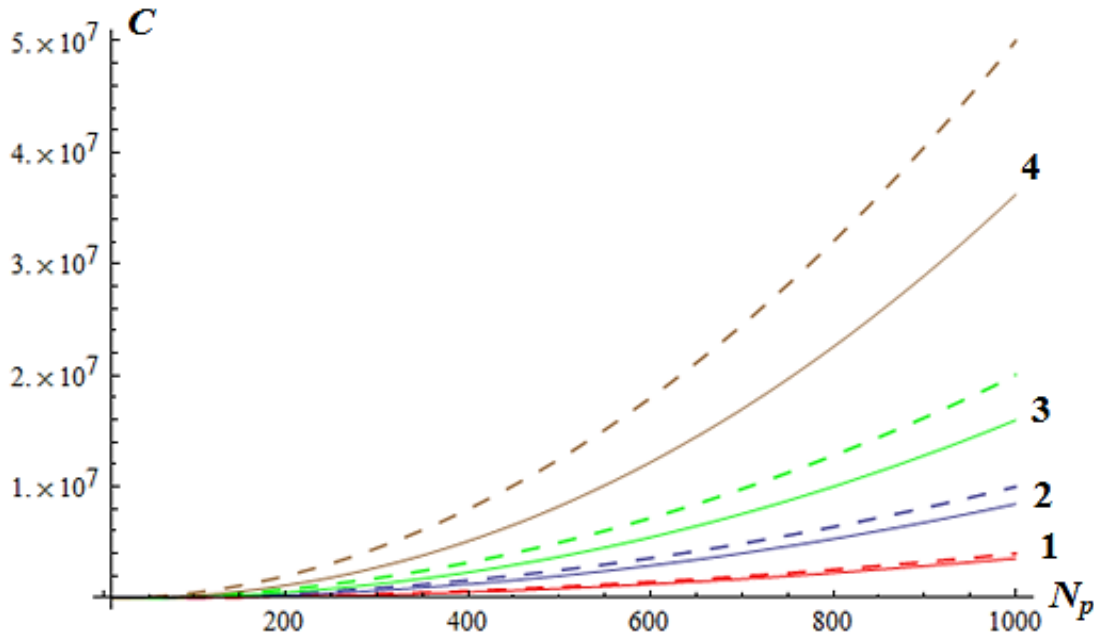


Рис. 4. Залежність складності навчання від кількості навчальних прикладів. 1 – $N_{in} = 4$; 2 – $N_{in} = 10$; 3 – $N_{in} = 20$; 4 – $N_{in} = 50$; штрихові лінії – лінійне наближення

Із означення параметра складності $C = N_p \times (N_{\lambda_{ij}}^{opt})^2$ випливає практичний висновок: для нейроммуляторів із продуктивністю сучасних персональних комп'ютерів ($P = 10^9$ операцій за секунду) з кількістю навчальних прикладів N_p і розмірністю входів N_{in} час навчання нейронної мережі визначатиметься таким співвідношенням $\tau = \frac{N_p \times (N_{\lambda_{ij}}^{opt})^2}{P}$. Зокрема, якщо $N_p = 500$, $N_{in} = 50$, $N_{\lambda_{ij}}^{opt} = 125$, час навчання нейронної мережі дорівнюватиме $\tau = 7,8 \cdot 10^{-3}$ с.

Висновки

Побудовано нелінійну модель узагальненої похибки для нейронної мережі прямого поширення. На основі цієї моделі визначено оптимальну кількість синаптичних зв'язків, оптимальну кількість нейронів у прихованому шарі та параметр складності навчання нейронної мережі (кількість операцій, виконаних нейронною мережею).

Встановлено, що зі зростанням кількості вхідних образів N_{in} та кількості навчальних прикладів N_p мінімум узагальненої похибки зміщується у бік більшої кількості синаптичних зв'язків.

Показано, що зі зростанням кількості сенсорних нейронів N_{in} (вхідних образів) зменшується оптимальна кількість нейронів у прихованому шарі тришарової нейронної мережі.

References

1. Lytvyn, V., Peleshchak, I., Peleshchak, R. (2017). The compression of the input images in neural network that using method diagonalization the matrices of synaptic weight connections. *2nd International Conference on Advanced Information and Communication Technologies (AICT)*, 66–70. Retrieved from: <https://doi.org/10.1109/AIACT.2017.8020067>
2. Lytvyn, V., Peleshchak, I., Peleshchak, R. (2017). Increase the speed of detection and recognition of computer attacks in combined diagonalized neural networks. *4th International Scientific-Practical Conference "Problems of infocommunications. Science and Technolohy"*, 152–155. Retrieved from: <https://doi.org/10.1109/INFOCOMMST.2017.8246370>.
3. Yezhov, A. A., Shumsky, S. A. (1998). *Neurocomputing and its applications in economics and business*. Moscow, 222.
4. Tariq, R. (2016). *Make Your Own Neural Network*. Kindle Edition, 222.
5. Charu C. Aggarwal (2018). *Neural Networks and Deep Learning: A Textbook*. Springer, 520. Retrieved from: <https://doi.org/10.1007/978-3-319-94463-0>
6. Tereykovsky, I. A. (2012). Optimization of the structure of the multilayer perceptron in systems of computer informaton protection. *Information protection*, 3, 36–40. Retrieved from: <https://doi.org/10.18372/2410-7840.14.3357>
7. Landi, G., Zampini, A. (2018). *Linear Algebra and Analytic Geometry for Physical Sciences*. Springer, 345. Retrieved from: <https://doi.org/10.1007/978-3-319-78361-1>
8. Heydarov, P. S. (2017). Neural network of direct distribution with the calculated parameters. *Information technology*, 7, 543–552.
9. Hecht-Nielsen, R. (1987). Kolmogorov's mapping neural network existence theorem. *IEEE First Annual Int. Conf. on Neural Networks*. San Diego, 3, 11–13.
10. Kulchin, Y. N. (2015). Processing of signals of the distributed fiber-optic network for recognition of dynamic images using neural networks. *Information technology*, 4, 312–318.

**CONSTRUCTION OF AN OPTIMIZED MULTILAYER NEURAL NETWORK
WITHIN A NONLINEAR MODEL OF GENERALIZED ERROR**

Roman Peleshchak¹, Vasyl Lytvyn², Ivan Peleshchak³, Victoria Vysotska⁴, Oksana Chernyak⁵

¹⁻⁴Lviv Polytechnic National University

⁵Drohobych Ivan Franko State Pedagogical University

¹rpeleshchak@ukr.net, ORCID: 0000-0002-0536-3252

²vasyl.v.lytvyn@lpnu.ua, ORCID: 0000-0002-9676-0180

³peleshchakivan@gmail.com, ORCID: 0000-0002-7481-8628

⁴victoria.a.vysotska@lpnu.ua, ORCID: 0000-0001-6417-3689

⁵oksana2895@gmail.com, ORCID: 0000-0001-6034-1016

© Peleshchak R., Lytvyn V., Peleshchak I., Vysotska V., Chernyak O., 2021

In this paper, we propose a method for optimizing the structure of a multilayer neural network based on minimizing nonlinear generalized error, which is based on the principle of minimum length of description. According to this principle, the generalized error is determined by the error in the description of the model and the error in the approximation of the data by the neural network in the nonlinear approximation. From the condition of minimizing the generalized network error, the expressions for calculating the optimal network size are given (the number of synaptic connections and the number of neurons in hidden layers). The graphic dependences of the generalized error of the network on the number of synaptic connections between the neurons with different values of input images and the fixed number of educational examples and the graphic dependences of the optimal number of synaptic connections from the number of educational examples with different values of the input images are constructed. The assessment of the degree of complexity of the training of the neural network is carried out on the basis of the ratio of the optimal number of synaptic connections between the neurons and the optimal number of neurons in the hidden layers.

Key words: neural network; generalized error; size optimization.