УДК 004.032.26
DOI https://doi.org/10.32782/2663-5941/2024.3.1/09

Гайда А.Ю.
Національний університет кораблебудування імені адмірала Макарова

Морозова Г.С.
Національний університет кораблебудування імені адмірала Макарова

МЕХАНІЗМ АДАПТИВНОГО НАВЧАННЯ ГЛИБОКИХ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

Глибоке навчання останнім часом набуває все більшої популярності як за рахунок прогресу в обчислювальній техніці, так і завдяки демонстрації все більш успішних і переконливих результатів у вирішенні складних задач. Сучасні штучні нейронні мережі (ШНМ), навчені на значних обсягах даних, здатні ідентифікувати та класифікувати складні образи і явища, а глибокі ШНМ (ГШНМ) також здатні розпізнавати внутрішні неявні закономірності та зв'язки, оцінювати можливості, робити прогнози та приймати рішення. Створення, навчання і використання таких мереж вимагає комплексного розв'язання задач збору і підготовки даних, вибору архітектури мережі і процедури навчання тощо. Однією з проблем навчання ШНМ є проблема зникаючого градієнту, що обмежує застосування ШНМ лише для вирішення задач, що або не вимагають глибокого навчання, або пов'язані з даними, що мають просторову або часову локальність.

У статті розглянута проблема зникаючого градієнту, що має місце при навчанні повно-зв'язних ГШНМ та представлені результати аналізу особливостей зникаючого градієнту за результатами яких зроблено спробу часткового подолання зазначеної проблеми шляхом модифікації алгоритму навчання мережі. Отримані результати дозволили створити та здійснити навчання повно-зв'язних ГШНМ з числом прихованих шарів від 5 до 15. Апробації результатів здійснена на академічних наборах даних. Отримані результати можуть бути застосовані при розробці моделей навчання ШНМ.

Ключові слова: *штучна нейронна мережа, глибока штучна нейронна мережа, повно-зв'язна штучна нейронна мережанавчання штучної нейронної мережі, зникаючий градієнт, вибуховий градієнт, «Ірриси Фішера», «Червоне вино», «Біле вино».*

Постановка проблеми. Глибоке навчання – це галузь машинного навчання, яка базується на глибоких штучних нейронних мережах (ГШНМ). Глибоке навчання дозволяє генерувати і використовувати складні моделі, досліджувати внутрішні залежності в даних і на їх основі продукувати нові знання. Не зважаючи на те, що штучні нейронні мережі здебільшого є програмними продуктами (хоча існують і цілком апаратні рішення), принципи їх роботи основані не на програмуванні, а на навчанні. Навчання здійснюється як направлений пошук передаточної функції мережі, щоб найбільш точно відтворити вихідні значення з вхідного набору даних (званого навчальною вибіркою) для відповідних вхідних значень з цього ж набору і, очікувано, знайти вихідні значення для даних, що не входять до навчальної вибірки. Результати навчання зберігаються як вагові коефіцієнти в структурі мережі і можуть розглядатись як зважені асоціації між вхідними даними та результатами часткової або повної обробки вхідних даних.

Не зважаючи на велику кількість бібліотек і пакетів програм для побудови ГШНМ, створення ефективної мережі, здатної при вирішенні складної задачі надавати надійні і достовірні результати, вимагає застосування, а іноді і пошуку певних правил збору і підготовки вхідних даних, вибору архітектури мережі, технологій навчання тощо, що відповідають складності задачі і вимогам до результатів.

Аналіз останніх досліджень і публікацій. Однією з нагальних проблем навчання ГШНМ є проблема зникаючого (вибухового) градієнту (vanishing gradient), яка полягає в тому, що значення помилки, отриманої на виході мережі, експоненціально зменшується при передачі від виходу до входу (якщо градієнт менше 1) або стрімко зростає (якщо градієнт більше 1). Одним із поширених засобів подолання цієї проблеми є зменшення швидкості навчання, що призводить до уповільнення процесу навчання, або, навіть, унеможливлює його [1–3]. Як відомо, проблема зникаючого градієнту є наслідком застосування сигмоїдальної передаточної функції нейрона для приведення зна-

чення виходу нейрона до визначеного діапазону (зазвичай від 0 до 1.0 або від –1.0 до 1.0). Найбільше ця проблема виявляється в повно-зв'язних штучних нейронних мережах. В згорткових ШНМ цю проблему вдалось частково подолати за рахунок заміни сигмоїдальної передаточної функції нейрона функцією обчислення згортки, що дозволило застосовувати лінійно-ламані функції активації. Але, впровадження згортки також обмежує область використання згорткових ШНМ лише задачами, що передбачають просторову або часову локальність даних – як, наприклад, зображення, звук, динаміка фінансових показників тощо. Таким чином, повно-звязні ШНМ здатні вирішувати задачі різної природи, але їх використання обмежене неможливістю застосування у випадках, що вимагають глибоких мереж, а згорткові ШНМ здатні вирішувати задачі, що вимагають глибоких мереж, але їх використання обмежене необхідністю просторової або часової локальності даних. Все це разом обумовлює інтерес до подальших досліджень повно-зв'язних ШНМ.

Постановка завдання. Мета роботи – зменшити швидкості навчання, що призводить до уповільнення процесу навчання, або, навіть, унеможливлює його, дану проблему вирішити за рахунок забезпечення деякої «оптимальної» швидкості навчання кожного шару.

Виклад основного матеріалу. В роботі зроблено спробу аналізу особливостей навчання повно-зв'язних глибоких штучних нейронних мереж і запропоновані зміни до процесу навчання класичної повно-зв'язної ШНМ, які дозволили частково вирішити задачу подолання зникаючого (вибухового) градієнту.

Для вирішення поставленої задачі були створені глибокі (від 5 до 15 шарів) повно-зв'язні ШНМ (рисунок 1) різної архітектури. Для тестування розроблених ГШНМ були використані академічні набори даних – «Іриси Фішера», «Червоне вино», «Біле вино» – і обрані архітектури мереж, що відповідають складності цих даних. В процесі тестування вирішувалась задача класифікації. Всі набори даних були попередньо нормалізовані, а набори даних «Червоне вино» і «Біле вино» також були збалансовані шляхом циклічного додавання у набір наявних даних з міноритарних класів, недостатньо представлених у вхідному наборі. Початковий розподіл даних за класами для наборів даних «Червоне вино» і «Біле вино» представлений на рисунках 2 і 3. Набір даних «Іриси Фішера» балансування не потребував, оскільки всі представлені у наборі класи збалансовані [6]. Використання настільки

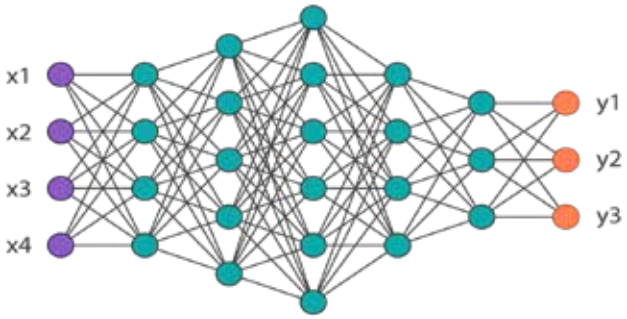

Рис. 1. Повно-зв'язна нейронна мережа

простого методу балансування для наборів даних «Червоне вино» і «Біле вино» обумовлене бажанням простого відтворення отриманих результатів та бажанням нівелювати вплив на отримані результати наявних методів балансування, що базуються на синтетичних даних – як, наприклад, SMOTE (Synthetic Minority Over-sampling Technique) [4].

В процесі навчання ГШНМ цілком виявилась проблема зникаючого градієнту і, як наслідок, мала місце низька якість навчання. Зважаючи, що ефект зникаючого градієнту виявляється як послідовне зменшення градієнту від виходу до входу на кожному з шарів мережі, була здійснена перевірка затухання градієнту на кожному з шарів мережі. Найбільш характерні значення градієнтів при моделюванні на зазначених наборах даних представлені на рисунках 4–6.

В результаті аналізу отриманих результатів було очікувано виявлено, що вхідні шари навчаються значно повільніше, ніж вихідні (що, загалом і складає проблему зникаючого градієнту), але також було виявлено, що ця проблема найбільше проявляється на пізніх стадіях навчання, а на початкових стадіях навчання основний внесок у помилку мережі

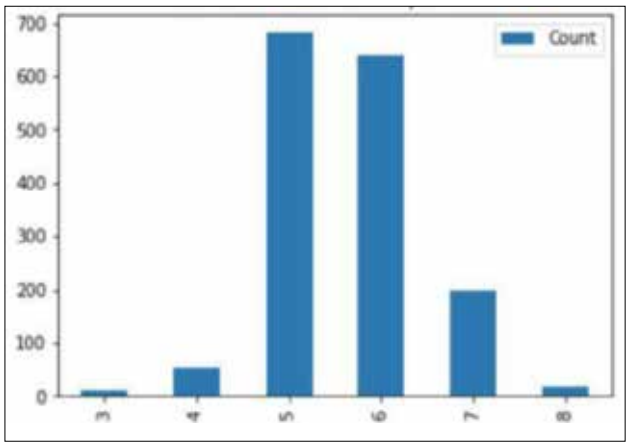

Рис. 2. Розподіл даних за класами для набору даних «Червоне вино»

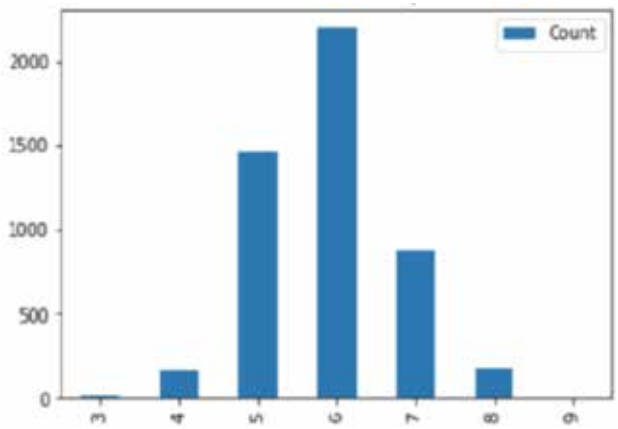

Рис. 3. Розподіл даних за класами для набору даних «Біле вино»

забезпечують не вхідні, а вихідні шари ГШНМ. Для зменшення впливу зазначеного ефекту була розроблена модель навчання, в якій швидкість навчання визначалась на кожному кроці навчання залежно від середнього градієнту шару, що дозволило збільшити швидкість навчання.

Розрахунок значення швидкості навчання $S_{i,k}$ для кроку навчання i з шару k здійснювався за формулою:

$$S_{i,k} = S_{i-1,k} + a_k \left(g_i G_{i,k} - S_{i-1,k} \right),$$

де a_k – коефіцієнт згладжування; g_i – коефіцієнт приведення градієнту до масштабу швидкості навчання; $G_{i,k}$ – середнє значення градієнту для кроку i шару k.

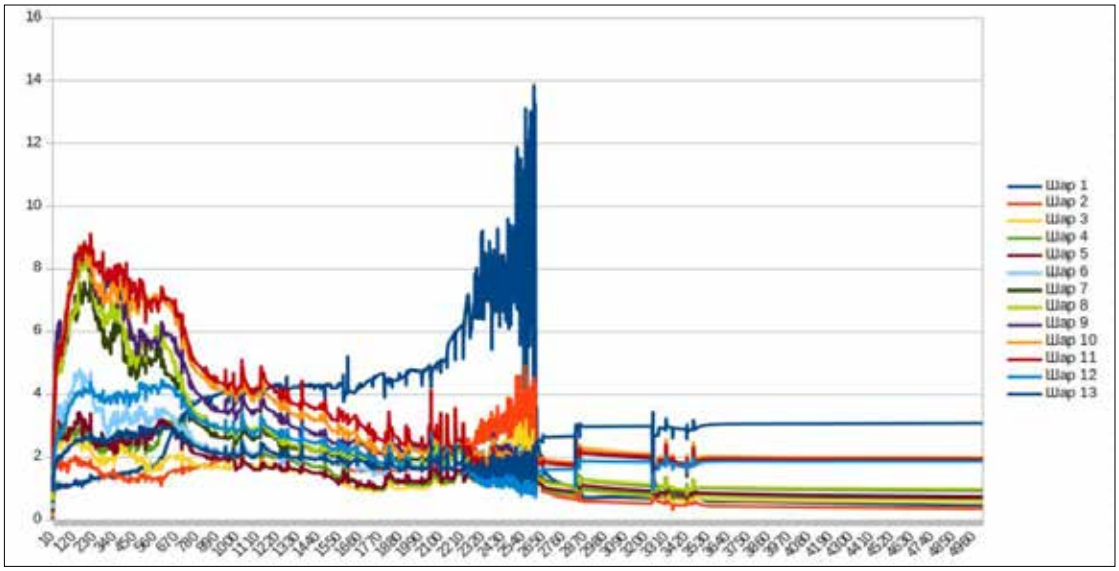

Рис. 4. Розподіл градієнтів даних за шарами при навчанні мережі на наборі даних «Іриси Фішера»

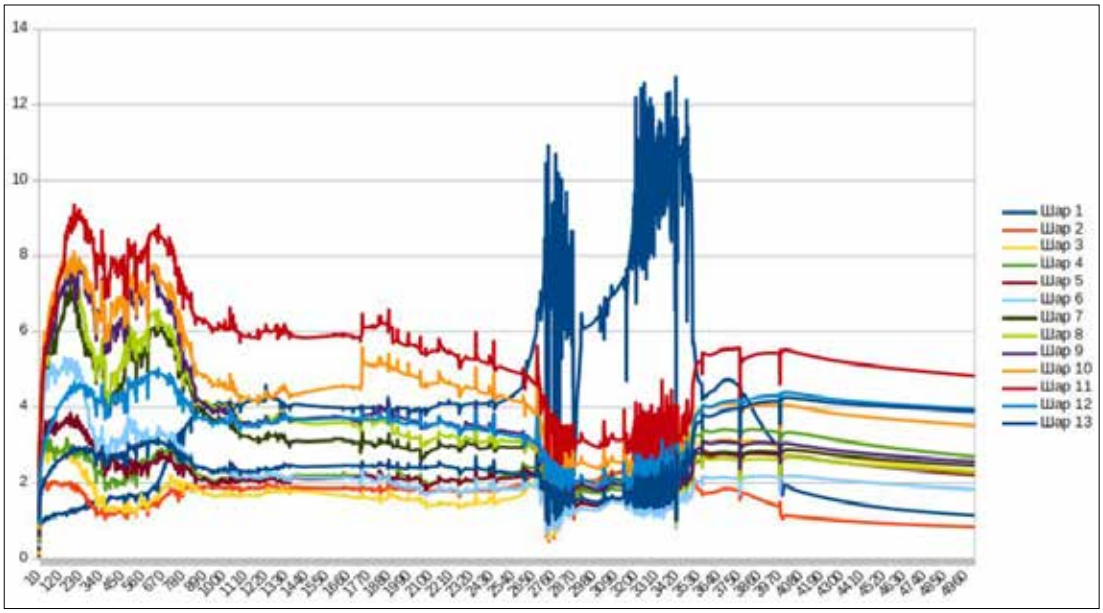

Рис. 5. Розподіл градієнтів даних за шарами при навчанні мережі на наборі даних «Червоне вино»

Встановлення швидкості навчання окремо для кожного шару ШНМ залежно від поточного градієнту шару, дозволило підвищити швидкість навчання і зменшити похибку мережі, але, у свою чергу, часто призводило до швидкого перенавчання мережі. Для запобігання ефекту перенавчання був застосований популярний засіб – алгоритм виключення (dropout).

Розрахунок значення ймовірності виключення $P_{i,k}$ нейрона з шару k для кроку навчання i здійснювався за формулою:

$$P_{i,k} = P\frac{M + G_{i,k}}{M + G_{i-1,k}},$$

де P – коефіцієнт виключення; M – стабілізувальна складова для зменшення викидів; $G_{i,k}$ – середнє значення градієнту для кроку i шару k.

На рисунку 7 представлені результати апробації розробленої моделі навчання ГШНМ в прикладній задачі. Як видно з рисунку, запропонована модель дозволила в 3–4 рази покращити точність отриманих результатів порівняно з традиційною моделлю навчання.

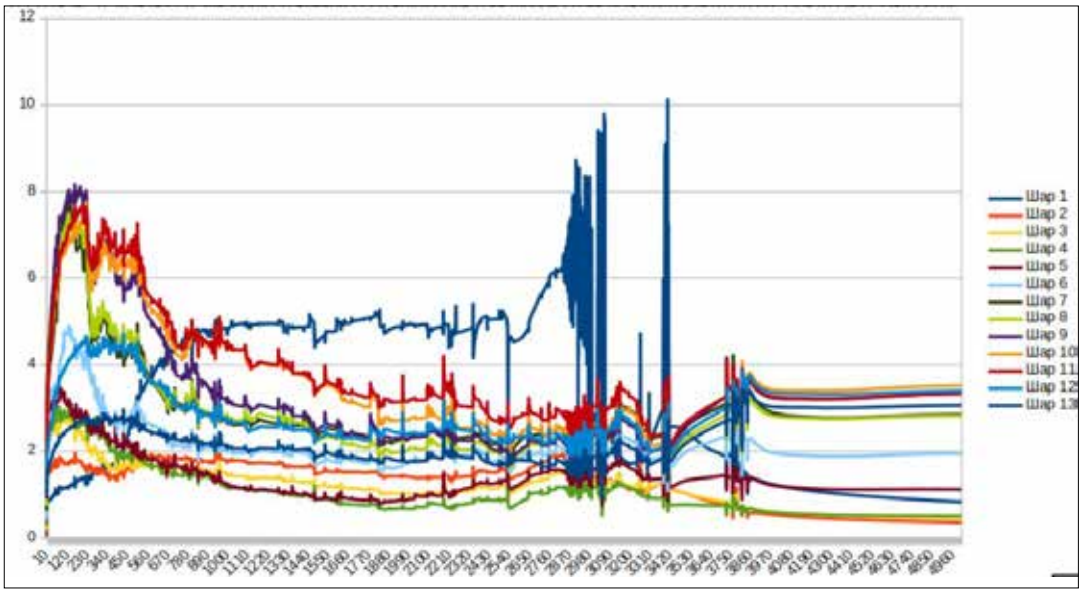

Рис. 6. Розподіл градієнтів даних за шарами при навчанні мережі на наборі даних «Біле вино»

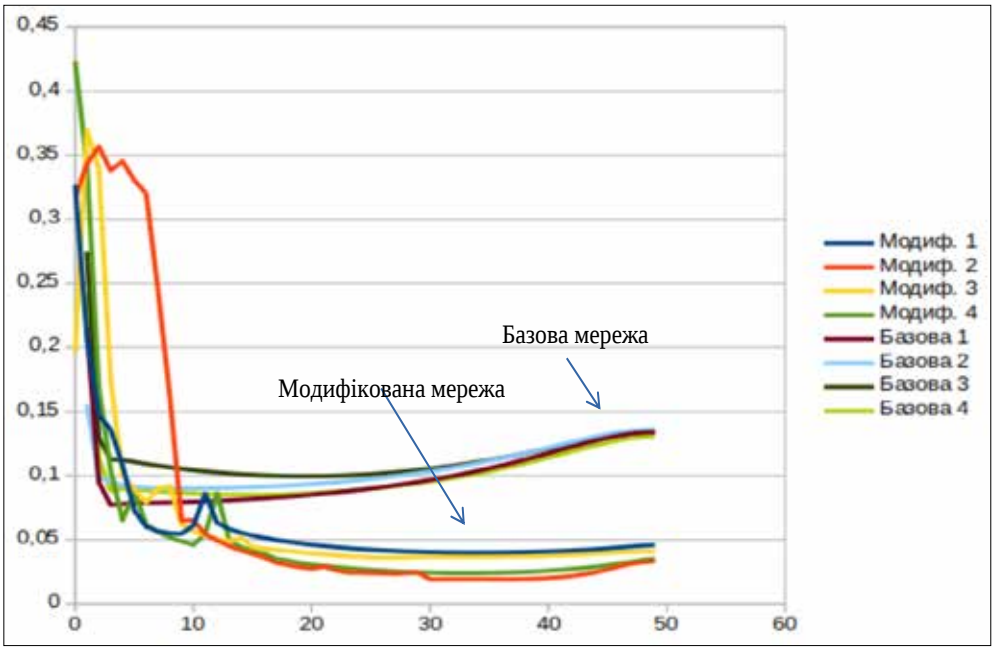

Рис. 7. Значення помилки базової та модифікованої мережі від числа кроків навчання (×100) для наборів даних

Висновки. На основі отриманих результатів моделювання можна зробити висновок, проблема зникаючого градієнту повно-зв'язних ГШНМ може бути частково вирішена за рахунок забезпечення деякої «оптимальної» швидкості навчання кожного шару. Побудова моделі «оптимальної» швидкості навчання ШНМ вимагає додаткових досліджень.

Список літератури:

1. Hochreiter, S. (1991). *Untersuchungen zu dynamischen neuronalen Netzen* (Diplom thesis). Institut f. Informatik, Technische Univ. Munich.

2. Hochreiter, S.; Bengio, Y.; Frasconi, P.; Schmidhuber, J. (2001). Gradient flow in recurrent nets: the difficulty of learning long-term dependencies / S. C. Kremer, J. F. Kolen. *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press. ISBN 0-7803-5369-2.

3. Goh, Garrett B.; Hodas, Nathan O.; Vishnu, Abhinav (15/07/2017). Deep learning for computational chemistry. *Journal of Computational Chemistry /* 1291–1307. Bibcode:2017arXiv170104503G. PMID 28272810. arXiv:1701.04503. doi:10.1002/jcc.24764.

4. Synthetic Minority Over-sampling Technique (SMOTE) for Predicting Software Build Outcomes / Russel Pears, Jacqui Finlay, Andy M. Connor / Cornell University https://arxiv.org/abs/1407.2330

5. Гайда А.Ю., Фаріонова Т.А., Ворона М.В. Механізми ефективного управління динамічними системами з нечітко вираженими станами / О.Ю. Гайда, Т.А. Фаріонова, М.В. Ворона. *Інновації у суднобудуванні та океанотехніці* : VII Міжнародна науково-практична конференція. Миколаїв, 2016.

6. Гайда А.Ю. Михелєв І.Л., Морозова Г.С. До проблеми зникаючого градієнту при навчанні глибоких штучних нейронних мереж. *Інновації в суднобудуванні та океанотехніці* : XIV Міжнародна науково-технічна конференція. Миколаїв, 2023.

Hayda A.Yu., Morozova H.S. MECHANISM FOR ADAPTIVE TRAINING OF DEEP ARTIFICIAL NEURAL NETWORKS

Deep learning has recently gained more and more popularity both due to progress in computing technology and due to the demonstration of increasingly successful and convincing results in solving complex problems. Modern artificial neural networks (ANNs), trained on large amounts of data, are able to identify and classify complex images and phenomena, and deep ANNs (ANNs) are also able to recognize internal implicit patterns and connections, evaluate opportunities, make predictions and make decisions. The creation, training and use of such networks requires a comprehensive solution to the problems of data collection and preparation, selection of network architecture and training procedures, etc. One of the problems of learning ANNs is the problem of vanishing gradient, which limits the application of ANNs only to solving problems that either do not require deep learning or are related to data that has spatial or temporal locality.

The article examines the problem of the vanishing gradient, which occurs during the training of fully connected GSNs, and presents the results of the analysis of the features of the vanishing gradient, based on the results of which an attempt was made to partially overcome the specified problem by modifying the network learning algorithm. The obtained results made it possible to create and implement the training of fully connected GSHNMs with the number of hidden layers from 5 to 15. The results were tested on academic data sets. The obtained results can be applied in the development of ANN training models.

Key words: *artificial neural network, deep artificial neural network, fully connected artificial neural network, artificial neural network learning, vanishing gradient, explosive gradient, "Fisher's irises", "Red wine", "White wine".*