

напряму, урахуванням потреб суднобудівних, машинобудівних та ІТ підприємств Миколаївщини. Це забезпечує можливість отримання якісної професійної освіти в галузі ІТ та відповідає вимогам ринку праці України.

Саме вимоги роботодавців та побажання випускників кафедри ІУСТ, які є професіоналами-практиками та працюють на провідних посадах в ІТ-компаніях України та за кордоном, обумовлюють перехід на вивчення найсучасніших технологій та засобів розробки. Зокрема, мова програмування Kotlin має більше використовуватись в рамках таких навчальних дисциплін, як "Алгоритмізація та програмування", "Web-технології та web-дизайн", "Технологія розробки мобільних застосунків".

Запропонований підхід щодо підготовки студентів ІТ-галузі відповідає теоретичному змісту предметної області, спрямований на нове покоління програмних продуктів, сучасні методології проектування інформаційних систем та системний інформаційний менеджмент.

Література

[1]. Наказ Міністерства науки і освіти України "Про затвердження Положення про акредитацію освітніх програм, за якими здійснюється підготовка здобувачів вищої освіти" за № 977 від 11.07.2019 р. [Електронний ресурс]. – Режим доступу: http://search.ligazakon.ua/l_doc2.nsf/link1/RE33851.html

[2]. Є.Ю. Беркунський, А.Ю. Павленко. Алгоритмізація та програмування мовами Kotlin, C/C++ навчальний посібник. – Миколаїв: НУК, 2022. – 256 с.

Modern development tools and programming languages in the educational process

Yevhen Berkunskyi, Alona Pavlenko, Ihor Mykhelev, National University of Shipbuilding

Abstract: The choice of programming languages and development tools within the model of training students of Admiral Makarov National Shipbuilding University in the field of information technologies with an emphasis on the technical direction and taking into account the needs of employers in accordance with the requirements of the labor market for IT specialists in Ukraine is substantiated.

Keywords: Kotlin, Java, Programming language, Framework, Modern design technologies, Accreditation of educational programs, Criteria for evaluating the quality of an educational program, Quality management system

УДК 004.43

ДО ПРОБЛЕМИ ЗНИКАЮЧОГО ГРАДІЄНТУ ПРИ НАВЧАННІ ГЛИБОКИХ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

Гайда А.Ю.¹, Михелєв І.Л.¹, Морозова Г.С.²

к.т.н., доцент кафедри інформаційних управляючих систем та технологій¹,

ст. викладач кафедри інформаційних управляючих систем та технологій²

Національного університету кораблебудування імені

адмірала Макарова

м. Миколаїв, Україна

cetus@ukr.net¹, mihelevigor@gmail.com¹, amorozova711004@gmail.com²

Анотація: У статті розглянута актуальна проблема зникаючого градієнту, що має місце при навчанні і використанні повно-зв'язних глибоких штучних нейронних мереж. У роботі зроблено спробу аналізу особливостей виникнення проблеми зникаючого градієнту глибоких

нейронних мереж та зроблено спробу часткового подолання зазначеної проблеми шляхом модифікації алгоритму навчання мережі. Показані результати практичної апробації результатів аналізу і розроблених рішень.

Ключові слова: штучна нейронна мережа, глибока штучна нейронна мережа, навчання штучної нейронної мережі, зникаючий градієнт, “Іриси Фішера”, “Червоне вино”, “Біле вино”.

Вступна частина.

Штучні нейронні мережі (ШНМ) — це категорія алгоритмів машинного навчання, що на основі наявних даних про процеси, образи, предмети, мову тощо дозволяють вирішувати задачі класифікації, розпізнавання, прогнозування, фільтрування, генерації та інші. При вирішенні окремих задач можливості ШНМ значно перевищують людські. Наприклад, нейромережа AlphaZero, навчена грі у шахи в процесі гри зі своєю копією, перемагає кращі шахові програми, окремим з яких, у свою чергу, поступалися гротмейстери і чемпіони світу.

Мета роботи полягає у розробці моделі навчання, яка на основі результатів аналізу особливостей навчання повно-зв'язних глибоких штучних нейронних мереж з алгоритмом зворотнього поширення помилки дозволить частково або повністю подолати проблему зникаючого (вибухового) градієнту.

Основна частина.

Вирішення окремих з таких задач потребує здійснення значної кількості послідовних узагальнень подібно тому, як людський мозок здійснює серію узагальнень на зображенні, що передається з відеокамери, яка віддаляється. Наприклад: спочатку пісок, потім доріжка з піску, потім доріжка з піску між дерев сприймається як парк. Загалом, порядок таких узагальнень може бути і іншим, але важливим є те, що для ефективного вирішення багатьох задач необхідні засоби, які дозволять ефективно здійснювати такі серії узагальнень. Одним з таких засобів є глибокі штучні нейронні мережі.

Глибокі штучні нейронні мережі (ГШНМ) — це категорія нейронних мереж, та, загалом, алгоритмів машинного навчання, реалізована шляхом послідовного укладання декількох шарів нейронної мережі таким чином, що кожен наступний шар здійснює обробку результатів, отриманих від попереднього шару і передає результати такої обробки як власний результат для подальшої обробки у наступний шар (можливі деякі відхилення від такої моделі, але принцип послідовної обробки тут є ключовим). Кількість шарів ШНМ, що дозволяє кваліфікувати таку мережу як ГШНМ на поточний момент чітко не визначена і це питання є дискусійним — в окремих джерелах кількість таких шарів визначається як “декілька”, в інших як “понад 3” або “7 і більше” (в окремих сучасних мережах число шарів ГШНМ може сягати понад 100). Тут треба зауважити, що при підрахунку таких шарів маються на увазі лише так звані приховані шари, а вхідний (через який подаються вхідні дані) і вихідний (з якого знімають результати обробки) шари не враховуються, оскільки такі шари присутні і в звичайній, “неглибокій” ШНМ (хоча вихідний шар також здійснює обробку даних і обчислює результат).

Незважаючи на велику кількість готових рішень для побудови ГШНМ (бібліотек та пакетів програм), створення ефективної мережі, здатної не лише виконувати обробку вхідних даних, але і надавати при цьому надійні і достовірні результати, є складною задачею, вирішення якої можливе лише за умов коректного збору і підготовки вхідних даних, коректного вибору архітектури мережі, коректної процедури навчання тощо. Під “коректними” тут мається на увазі такі збір і навчання, вхідні дані і архітектура тощо, які відповідають природі і складності задачі, вимогам до результатів обробки тощо. Досягнення окремих умов забезпечення такої коректності знаходиться на межі між науковими методами та досвідом і особистим розумінням задачі конструктором мережі. Але отримані результати варті додаткових зусиль і про це свідчить вибух інтересу до штучного інтелекту загалом і ГШНМ зокрема.

Однією з серйозних проблем побудови ГШНМ є так звана проблема зникаючого (а також 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), що не дозволяє виконати навчання ШНМ [1-3]. Часткове вирішення цієї проблеми, отримане в результаті створення згорткових ШНМ (ЗШНМ) стало знаковою подією, але ЗШНМ дозволяють вирішувати лише такі задачі, в яких вхідні дані відбивають певну просторову або часову організацію, наприклад, аналіз зображень (просторова організація) або звуку (часова організація) [5].

В роботі зроблено спробу аналізу особливостей навчання повно-зв'язних глибоких нейронних мереж і проблем, що виникають при цьому, та розробки моделі, яка дозволить частково або повністю подолати проблему зникаючого (вибухового) градієнту.

Для вирішення поставленої задачі авторами були розроблені глибокі (15 шарів) ШНМ різної архітектури. Для тестування розроблених повно-зв'язних ГШНМ були використані академічні набори даних — “Іриси Фішера”, “Червоне вино”, “Біле вино” — і обрані відповідно складності цих даних архітектури мереж. 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 представлені у вхідному наборі. Для вин початковий розподіл даних за класами представлений на рисунках 1 і 2. Набір даних “Іриси Фішера” балансування не потребував, оскільки всі класи представлені у наборі збалансовані. Використання настільки простого методу балансування наборів даних обумовлене бажанням нівелювати можливий вплив на отримані результати 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 і отриманих результатів. Як було зазначено вище, такі результати цілком очікувані і причина добре відома — зникаючий градієнт. Зважаючи, що ефект зникаючого градієнту виявляється як послідовне зменшення градієнту на кожному з шарів мережі, була здійснена перевірка затухання на кожному з шарів мережі. Найбільш характерні результати моделювання для градієнтів на зазначених наборах даних представлені на рисунках 3-5.

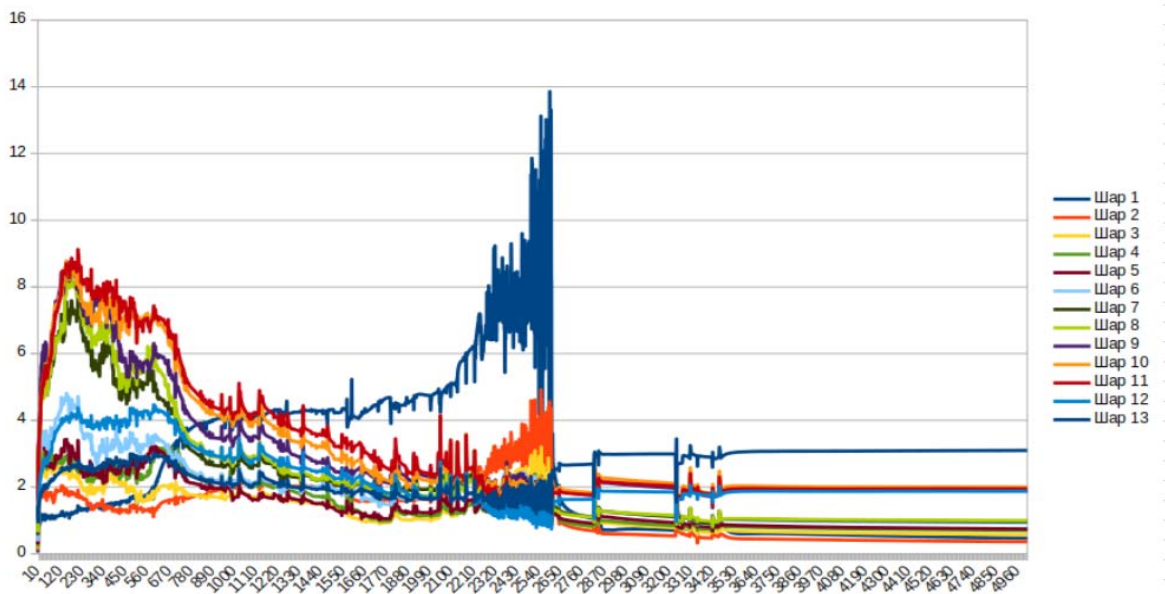


Рис. 1. Розподіл відношення помилки шару до помилки мережі для набору даних “Іриси Фішера”

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 забезпечують вихідні шари ГШНМ.

Для зменшення впливу зазначеної проблеми на процес навчання мережі була розроблена адаптивна модель навчання, яка встановлює швидкість навчання (отже і градієнт) для кожного шару ШНМ залежно від середнього градієнту шару.

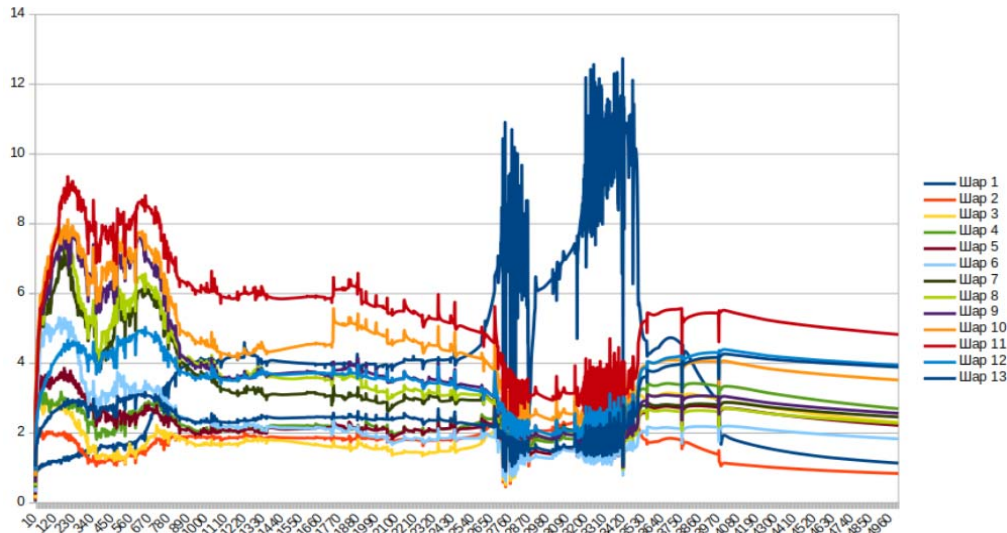


Рис. 2. Розподіл відношення помилки шару для набору даних “Червоне вино”

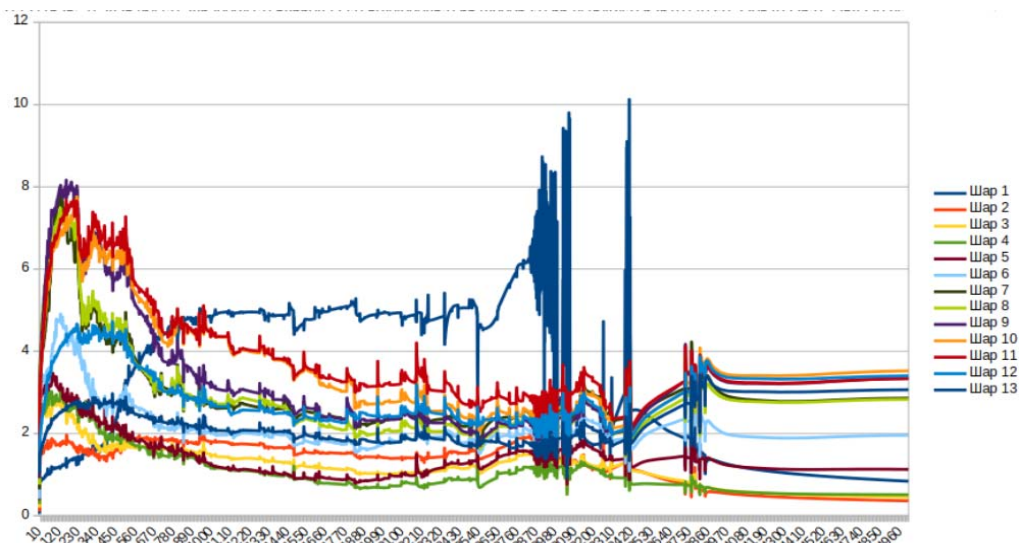


Рис. 3. Розподіл відношення помилки шару до помилки мережі для набору даних “Біле вино”

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

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

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 .

На рисунку 4 представлені результати апробації на прикладній задачі модифікованої мережі на основі розробленої моделі навчання ГШНМ. З результатів апробації видно, що запропонована модель має в 3-4 рази меншу помилку навчання, що порівняно з традиційною моделлю навчання дозволило значно покращити точність отриманих результатів.

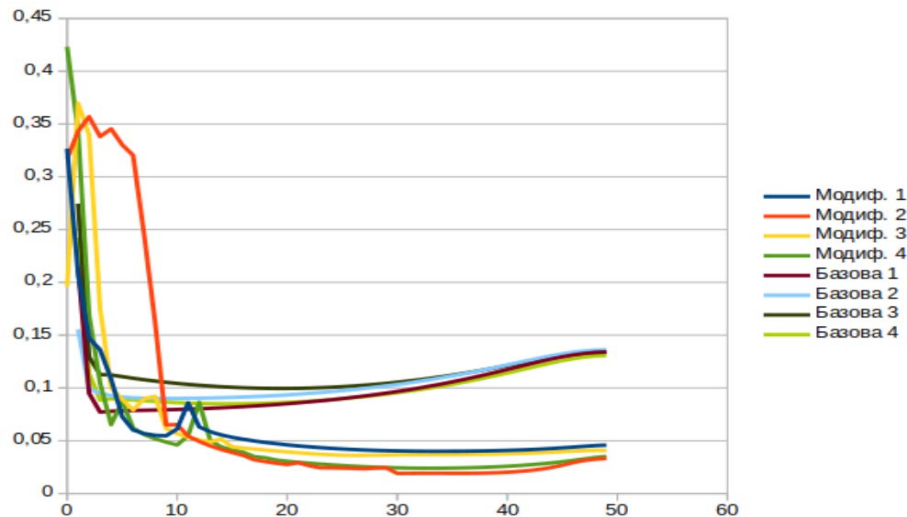


Рис. 4. Значення помилки базової та модифікованої мережі від числа кроків навчання (x100)

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.

Література

- [1]. S. Hochreiter (1991). Untersuchungen zu dynamischen neuronalen Netzen (Diplom thesis). Institut f. Informatik, Technische Univ. Munich.
- [2]. S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber, (2001). Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. / Kremer, S. C.; Kolen, J. F. A Field Guide to Dynamical Recurrent Neural Networks. IEEE Press. ISBN 0-7803-5369-2.
- [3]. Goh, B. Garrett, 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]. А.Ю. Гайда, Т.А. Фарионова, М.В. Ворона. Механизмы эффективного управления динамическими системами с нечетко выраженными состояниями / А.Ю. Гайда, Т.А. Фарионова, М.В. Ворона / "Інновації в суднобудуванні та океанотехніці". VII Міжнародна науково-практична конференція – Миколаїв: 2016.

Towards the problem of vanishing gradient in learning deep artificial neural networks

Anatolii Gaida, Ihor Mykhelev, Hanna Morozova, National University of Shipbuilding

Abstract: The article discusses the actual problem of the vanishing gradient, which occurs when training and using fully connected deep artificial neural networks. In the work, an attempt was made to analyze the peculiarities of the problem of the vanishing gradient of deep neural networks and an attempt was made to partially overcome the specified problem by modifying the network learning algorithm. The results of the practical testing of the analysis results and developed solutions are shown.

Keywords: artificial neural network, deep artificial neural network, learning of artificial neural network, vanishing gradient, "Fisher irises", "Red wine", "White wine".