

УДК 004.032.26:681.883

DOI: 10.31673/2412-9070.2023.039000

А. О. ОЛЕКСІЙ, аспірант;

А. А. ВЕРЛЯНЬ, канд. техн. наук, доцент,

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ

ЗАСТОСУВАННЯ БАГАТОШАРОВОГО ПЕРСЕПТРОНА ДЛЯ АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ У ВОДНОМУ СЕРЕДОВИЩІ

Задача аналізу акустичних сигналів водного середовища є досить непростою. Записані у складному середовищі дані можуть страждати від низьких відношень сигналу та шуму. Тож застосування нейромереж імовірно стане ефективним підходом, але навіть за наявності задовільних результатів удосконалення методів сприятиме подальшому підвищенню точності класифікації, забезпечуючи більш надійні та точні результати за умов високої шумової активності. Є потреба у створенні поліпшеного нейромережного методу для класифікації шумів водного середовища з різним відношенням сигналу та шуму. Удосконалення досягається через упровадження багатошарового персептрона з адаптивним навчанням та пакетною нормалізацією. Запропонований підхід дає змогу ефективно аналізувати шуми водних суден за умов високого рівня шуму та у складному середовищі. У результаті вдалось покращити метод та досягти точності класифікації в 96%, тоді як у разі застосування оригінального методу максимальна точність класифікації становила 94%.

Ключові слова: машинне навчання; акустичні сигнали; підводне середовище; класифікація.

Вступ

Постановка проблеми. Завдання аналізу акустичних сигналів водного середовища є складною задачею, яка може полягати в низькому відношенні сигналу та шуму або малій кількості даних. Одним із найбільш ефективних способів вирішення цих проблем є нейромережні підходи. Сьогодні задача аналізу акустичних сигналів морського середовища має багато складних підзадач, серед яких можуть бути такі проблеми: незначна якість одержуваної інформації та висока швидкість зміни навколишнього середовища. Під час пасивних методів приймання сигнали можуть характеризуватись невисоким відношенням сигнал/шум.

Для задачі класифікації акустичних сигналів водного середовища існує кілька основних видів нейромережних архітектур, що можуть бути застосовні. До них належать такі архітектури: згорткові нейронні мережі (*Convolutional Neural Networks, CNN*), рекурентні нейронні мережі (*Recurrent Neural Networks, RNN*), багатошарові персептрони (*Multi-layered Perceptron, MLP*), автоенкодер (*Autoencoder, AE*) тощо.

Багатошаровий персептрон здатний виявляти складні та неочевидні зв'язки між акустичними характеристиками сигналів та їх класами. Це дає змогу моделі вчитись на основі більш високорівневих особливостей. MLP може працювати з різними видами акустичного подання, включно зі спектрограмами, амплітудами та іншими часово-частотними характеристиками. Його застосування з додатковими шарами дає можливість моделі навчатись глибоким та ієрархічним поданням даних, що може бути корисним для роз-

пізнавання багатошарових акустичних шаблонів. Автоенкодер може застосовуватись для відбору важливих акустичних характеристик із великої кількості даних. Енкодер буде вивчати компресовані подання сигналів, що можуть бути використані для подальшої класифікації, а декодер допоможе відтворити сигнали для аналізу та інтерпретації. Підходи з використанням АЕ можуть допомогти виявляти приховані акустичні шаблони та корисні особливості у сигналах, що можуть бути потрібні для задачі класифікації.

Також можливі гібридні нейромережі, де задачу класифікації виконуватиме інший шар або інша нейромережа. Моделі, створені на базі CNN, особливо ефективні для оброблення зображень та відео. Вони використовують згорткові шари для виявлення локальних особливостей у вхідних даних. Під час їх застосування вхідні акустичні сигнали можуть бути перетворені як спектрограми, що дасть змогу CNN виявляти певні акустичні шаблони та залежності.

Якщо акустичні сигнали мають деяку послідовність або часову залежність, то можуть бути застосовні RNN. Вони мають здатність запам'ятовувати попередні стани та використовувати їх для оброблення нових вхідних даних. Мережі довготривалої короткочасної пам'яті (*Long Short-Term Memory, LSTM*) — це спеціальні типи RNN, які допомагають уникнути проблеми загасання градієнта та дають можливість більш ефективно працювати з довгими послідовностями даних. Далі буде виконано огляд літератури, де нейронні мережі застосовуються для задач аналізу акустичних сигналів водного середовища.

© А. О. Олексій, А. А. Верлянь, 2023

Аналіз останніх досліджень і публікацій. У статті [1] було запропоновано нейромережну архітектуру акустичного розпізнавання підводного сигналу, засновану на густій згортковій нейромережі (*Underwater Acoustic Target Classification Based on Dense Convolutional Neural Network, UATC-DensNet*). Архітектуру нейромережі розроблено з метою повторного використання попередніх карт характеристик, що дає змогу уникнути змивання градієнта. Наявність пропускових зв'язків уможливорює повторне використання всіх попередніх карт ознак, запобігаючи загасанню градієнта, спричиненому накладанням багатьох послідовних шарів згортки та активації. На вхід подається представлення сигналу в часовому діапазоні. Запропонований метод було протестовано на даних, записаних за умов реального підводного середовища з різним відношенням сигналу та шуму (*signal-to-noise ratio, SNR*). Цей метод показав точність класифікації на рівні 98,85%.

У [2] було досліджено використання мережі CNN разом із методом покращення спектра низькочастотного аналізу та записів (*Low-Frequency Analysis and Recording, LOFAR*). Для вирішення проблеми існування розривів у спектрі LOFAR, що може виникнути за наявності численних шумів або високого відношення сигнал/шум, було запропоновано метод підсилення спектра LOFAR на основі багатоступеневого алгоритму ухвалення рішень у спектрі LOFAR. Зі спектра LOFAR підводного акустичного сигналу, який здобуто багатоступеневим аналізом, розробляється структура згорткової нейронної мережі відповідно до його характеристик. Для цієї структури в CNN застосовуються деякі ідеї початкового модуля, який використовує різні розміри ядер згортки та вагує ознаки глобального та локального розподілів інформації. Ця структура мережі вибирає різні ядра згортки та ядра групування для попереднього вилучення ознак. Експериментальні дані, використані в цій статті, було поділено на дві частини. Результати показали, що метод LOFAR-CNN може досягти найвищої точності розпізнавання на рівні 95,22%, ще більше підвищуючи точність порівняно з іншими традиційними методами. Цей підхід дає змогу класифікувати дані з різними значеннями відношення SNR та за різних умов навколишнього середовища.

У праці [3] розглянуто новий метод, в якому інтегровано згорткову нейромережу та нейромережу довготривалої короткочасної пам'яті. Згорткова нейромережа використовує одновимірну згортку для оброблення вхідних даних. Оскільки характеристики акустичного сигналу підводного судна змінюються в часі, можна використовувати LSTM для роботи з характеристиками поточного моменту та збереженою інформацією про попередній

момент. У поєднанні з одновимірною згортковою нейромережею та мережею LSTM система може швидко адаптуватися до змін сигналу та підвищити точність розпізнавання. Частина згорткової нейромережі складається з двох шарів згортки та двох шарів злиття по черзі. Шар злиття використовує максимальне злиття, за яким іде шар відкидання. Нейромережа LSTM складається з одного шару LSTM та одного шару відкидання. Зрештою, вона подається в повнозв'язний шар для класифікації. Експериментальні результати із використанням набору даних підводних суден показали середню точність 92%, що на 8% вища, ніж CNN, і на 15% вища за LSTM.

У статті [4] було запропоновано нову систему класифікації, яка об'єднує архітектуру CNN з групуванням другого порядку (*Second order pooling, SOP*) для видобутку часових кореляцій із часово-частотного подання випромінюваних акустичних сигналів. Згорткові шари застосовуються для навчання локальних характеристик за допомогою набору ядерних фільтрів із входів, які видобуваються завдяки константному Q-перетворенню (*Constant-Q transform, CQT*). Замість використання максимального групування, запропонований оператор SOP призначений для вивчення відповідей різних фільтрів CNN, використовуючи часову траєкторію ознак CNN для кожного частотного піддіапазону. Щоб зберегти частотні відмінності, корельовані ознаки кожної частоти зберігаються. Результати об'єднання нормалізуються за допомогою підписаного квадратного кореня та l2-нормалізації, а потім передаються на Softmax-класифікатор. Роботу мережі було перевірено на різних глибинах. Результати порівняли з результатами DAE та DBN. Вони засвідчили, що запропонована нейромережа продемонструвала на 8% вищі результати класифікації.

У [5] було розглянуто кооперативний метод глибокого навчання. Цей метод поєднує глибоку мережу короткочасної та довготривалої пам'яті (*Deep Long-short term memory network, DLSTM*) і глибокий автоенкодер (*Deep autoencoder, DAE*). Спочатку використовується автоенкодер LSTM, який навчає шари DLSTM. На другому етапі натреновані моделі DLSTM та класифікатора Softmax використовуються для класифікації підводних шумів. Перевірку запропонованого методу було проведено за допомогою набору даних, що містить реальні підводні звуки пасажирських та вантажних суден, а також навколишнього підводного шуму. Робота запропонованого методу порівнювалася з результатами застосування DAE, DLSTM і показала найвищу точність класифікації, яка становила 90%.

У статті [6] автори пропонують модель складеного розрідженого автоенкодера (*Stacked sparse autoencoder, SSAE*), яка становить низку прихова-

них шарів розріджених автоенкодерів та класифікатора Softmax. Розріджений автоенкодер здатен вивчити розріджене подання з вихідних даних. Для навчання більш дискримінативного подання, запропоновано побудувати модель SSAE з багатошаровою структурою, яка будується шар за шаром через об'єднання вхідного шару та прихованого шару розрідженого АЕ. Одночасно шар декодера для кожного окремого розрідженого АЕ вилучається. На вхід подавалися дані зі спектрального та вейвлет-діапазонів. Результати порівнювалися з результатами роботи DBN та SVM, де запропонований метод показав найвищу точність класифікації, яка становила 92%. Цей підхід виявляє стійкість до варіацій шуму та допомагає поліпшити узагальнювальні властивості нейромережі та дістати хороші результати за різних відношень сигнал/шум.

У праці [7] було проаналізовано метод мережі роздільного згорткового автоенкодера (*Separable convolution Autoencoder, SCAE*). Запропонована модель складається з мережі кодера та мережі декодера. У цьому дослідженні модифіковану версію Хсертіон було використано як основу для модуля кодера. Головна ідея мережі кодера на основі роздільної згортки полягає в тому, що відображення міжканальних та просторових кореляцій на картах ознак розділені. Використання цієї стратегії допомагає зробити навчання мережі більш ефективним та підвищити продуктивність з погляду точності класифікації. Роздільна згортка базується на ідеї роздільної згортки за глибиною, яка є просторовою згорткою, що виконується незалежно над кожним каналом вхідного сигналу. Вхідні дані подано характеристиками часо-частотного діапазону як зображення, серед яких: кепстр, мел-спектрограма, мел-частотні кепстральні коефіцієнти, гамматон-частотні спектральні коефіцієнти, вейвлет-пакети. Метод показав точність класифікації на рівні 77% під час роботи з CQT. Загалом, нейромережу можна використовувати для класифікації дуже великих та різноманітних наборів сигналів реального середовища, що дає відносно точні результати.

Формулювання мети статті. Метою цього дослідження є підвищення точності класифікації акустичних шумів водного середовища завдяки поліпшенню нейромережного підходу для аналізу акустичних сигналів водного середовища. Для цього пропонується дослідити застосування архітектури багатошарового перцептрона, адаптивне навчання та пакетну нормалізацію даних.

Основна частина

Підхід до розв'язання задачі було взято зі статті [6]. Далі буде подано попереднє оброблення вхід-

них сигналів, опис нейронної мережі, обґрунтування вибраних рішень та опис ходу експерименту.

Зазвичай спектральні характеристики використовуються завдяки їх легкості отримання та інформаційній насиченості; низькочастотні компоненти, що з'являються внаслідок обертання гвинтової лопатки, дають цінну інформацію для класифікації суден, особливо на великій відстані. Щоб дістати характеристику частоти валу (*shaft frequency feature, SFF*), аналізується спектральна густина потужності (*power spectrum density, PSD*) необроблених акустичних сигналів під водою. Оцінювання PSD містить неперервні спектральні компоненти, які можуть призвести до неточних результатів, якщо їх використовувати безпосередньо. Для усунення цього недоліку застосовується метод найменших квадратів поліноміального наближення (LSPF) $P_n(f_i)$, який вилучає неперервні спектральні компоненти $P_n(f_i)$ з PSD. Застосування LSPF мінімізує функцію витрат для відшукування оптимальних параметрів:

$$P_n(f_i) = \sum_{i=1}^N b_i f_i^i, \quad (1)$$

де $b_i \{f = 1, 2, \dots, n\}$ — параметри фільтрації P_n .

SFF можна дістати через віднімання апроксимаційного полінома від необробленої PSD:

$$P_i(f_i) = P(f_i) - P_n(f_i), \quad (2)$$

де $P_i(f_i) \{f_i = f_1, f_2, \dots, f_N\}$ — характеристика частоти валу.

PSD випромінюваного шуму об'єктів водного середовища містить цінну інформацію для класифікації. Постійний спектр суден, спричинений кавітацією гвинтової лопатки, відображає різноманітні профілі, що відтворюють структуру гвинтової лопатки та системи живлення. Складне підводне середовище та реверберація акустичних сигналів потребують адаптивного вилучення ознак. Глибокі мережі виявляють потенціал у поданні складних моделей, використовуючи глибоке навчання для поетапного виокремлення ефективних, вхідно-незмінних ознак. Характеристику LPS із вибраним піддіапазоном від 100 Гц до 1 кГц можна обчислити за формулою

$$s = \{s_1, s_2, \dots, s_p\}, \quad (3)$$

де p — розмірність характеристик LPS.

Пакетне вейвлет-перетворення (*Wavelet packet transform, WPT*) надає повну декомпозицію для аналізу нерівномірних акустичних сигналів під водою, на відміну від методів віконного перетворення Фур'є (*Short-time forier transform, STFT*). Енергія компонентів вейвлет-пакету (*Wavelet packet component energy, WPCE*) доводить свою ефективність у виявленні особливостей акустичних сигналів під водою в діапазоні часу та частоти. Цей підхід передбачає обчислення глобального розподілу енергії хвильового пакету для захоплення високо-

частотних миттєвих компонентів випромінюваних акустичних сигналів і визначається за рівністю

$$E_i^j = \frac{\sum_{k=1}^M \|d_j^i[k]\|^2}{\sum_{j=0}^{2^i-1} \sum_{k=1}^M \|d_j^i[k]\|^2}. \quad (4)$$

Після WPT необроблена послідовність сигналу перетворюється на послідовності піддіапазонів $d_j^i[k]$, де i позначає рівні декомпозиції, j — піддіапазони. Загальна енергія сигналу агрегується з використанням компонентних енергій хвильового пакету, визначаючи j -ту енергію піддіапазону i -го рівня декомпозиції E_i^j як суму квадратів амплітуд сигналів $d_j^i[k]$. Фокус зосереджено на піддіапазонах понад 1 кГц завдяки докладній височастотній інформації від WPT. Вибрану характеристичну ознаку WPCE T можна подати у вигляді

$$T = \{E_i^{B_1}, E_i^{B_2}, \dots, E_i^{B_g}\}, \quad (5)$$

де B_1, B_2, B_g — піддіапазони понад 1 кГц у діапазоні аналізу. При цьому, для пригнічення шумів застосовують низькочастотний та ширококутовий фільтри. Далі виокремлені ознаки комбінуються у вектори, які подаються для кожного з класів суден. Як результат, на вхід подавалися комбіновані дані зі спектральної та вейвлет-ділянок. Схему роботи попереднього оброблення зображено на рисунку.



Схема попереднього оброблення

Для класифікації корабельних шумів пропонується модель багатошарового перцептрона, що складається зі вхідного шару, кількох прихованих шарів і класифікатора Softmax. Для вивчення більш дискримінаційного подання запропоновано побудувати модель із багатошаровою структурою, яка формується через об'єднання вхідного і прихованого шарів. Вхідний шар містить 500 нейронів та до виходу застосовується функція активації ReLU. Після вхідного шару йдуть три приховані шари розмірністю 100 нейронів кожен. Для кожного з виходів нейронів прихованих шарів застосовується функція активації ReLU. До вхідного та першого прихованого шару було застосовано пакетну нормалізацію. Вихідний шар має чотири нейрони, кожен з яких відповідає класу суден. Як функція активації було взято функцію LogSoftmax. Також запропоновано застосувати

пакетну нормалізацію та адаптивне навчання для підвищення точності класифікації. Архітектуру нейромережі наведено в таблиці.

Архітектура нейромережі

Шар	Кількість шарів	Функція активації	Пакетна нормалізація
Вхідний шар	500	ReLU	+
Прихований шар	100	ReLU	+
Прихований шар	100	ReLU	-
Прихований шар	100	ReLU	-
Вихідний шар	4	Sigmoid	-

Пакетна нормалізація допомагає підтримувати постійний розподіл, зменшуючи ризик зіткнення з труднощами оптимізації та нестабільного навчання. Пакетна нормалізація дає можливість використовувати вищі темпи навчання. Нормалізовані активації зменшують імовірність вибуху або загасання градієнтів, даючи змогу більш ефективної оптимізації та зумовлює швидший збіг та кращий розв'язок. З огляду на застосування функції ReLU результат може бути кращим із використанням пакетної нормалізації, оскільки вони в обох випадках допускають як додатні, так і нульові значення. Також така комбінація може поліпши-

ти потік градієнтів через мережу, полегшуючи навчання змістовних ознак та подань. Пакетна нормалізація допомагає підвищити швидкість навчання та стабільність збіжності, дозволяючи легше обробляти градієнти та уникати проблем із зникненням/вибуванням градієнтів. У такому разі багатошаровий перцептрон може навчатись більш ефективно та швидко. Спрощення архітектури автоенкодера до багатошарового перцептрона може зменшити кількість параметрів моделі та складність навчання, що може знизити ризик перенавчання та забезпечити кращу загальну роботу моделі на нових даних.

Адаптивне навчання в нейромережах належить до використання таких технік, які динамічно налаштовують швидкість навчання та інші параметри під час процесу оптимізації. Метою адаптивного навчання є покращення процесу навчання,

прискорення збігу та підвищення ефективності навчання нейромережі. Методи адаптивного навчання спрямовано на подолання обмежень фіксованих темпів навчання і здатні прискорити збіг та покращення ефективності навчання нейромережі. Зменшення швидкості навчання допомагає процесу оптимізації подолати локальні мінімуми та дає змогу оптимізатору робити менші кроки, коли прогрес сповільнюється, що може дати кращі результати.

Адаптивні методи навчання надають більшу гнучкість навчанню нейромережі, змінюючи швидкість навчання або поведінку під час оптимізації протягом усього процесу навчання. Ці методи можуть привести до швидшого збігу, кращої оптимізації та більшого узагальнення. Порівняно з оригінальним методом багатошаровий перцептрон з адаптивним навчанням може дозволити побудувати більш просту та оптимізовану мережу для поставленої задачі. Використання адаптивного навчання дає змогу моделі більш гнучко адаптуватись до змін у навчальних даних або оточенні. Це може бути особливо корисним у разі існування динамічних змін у вхідних сигналах.

Висновки

У дослідженні було покращено метод класифікації акустичних сигналів водного середовища. Застосування пакетної нормалізації та адаптивного навчання дало змогу підвищити точність класифікації для різних класів суден та перевершити оригінальний метод.

Було створено датасет на основі датасету ShipsEar [8]. Ці сигнали складаються із записів, які містять чотири типи кораблів та поділяються на чотири класи. Файли було сегментовано на фрагменти розміром 2 с кожен. Усі класи мали файли з різним значенням SNR. Датасет розділено на набори для навчання, валідації та тестування у співвідношенні 72% до 20% і 8%. Також було реалізовано оригінальний метод SSAE. Під час порівняння двох методів багатошаровий перцептрон показав точність класифікації на рівні 96%, тоді як оригінальний метод показав точність класифікації на рівні 94%.

Проте точність класифікації не для всіх класів суден була високою. Один із класів мав у своєму

складі записи з дуже низьким відношенням SNR та низькою інтенсивністю сигналів. Точність класифікації для цього класу була нижчою порівняно з іншими класами. Наявність великої кількості даних такої якості може призвести до погіршення результатів. Тому є сенс працювати над вирішенням цих проблем. Можливим рішенням може стати розроблення окремої нейромережі для пригнічення шумів та її інтеграція із запропонованим класифікатором. Також можливим подальшим ходом дослідження є підвищення генералізаційних властивостей мережі завдяки тренуванню на різних вибірках сигналів водного середовища.

Список використаної літератури

1. Doan V.-S., Huynh-The T., Kim D.-S. Underwater acoustic target classification based on dense convolutional neural network // *IEEE Geosci. Remote Sens. Lett.*, Oct. 2020.
2. Underwater target recognition based on multi-decision lofar spectrum enhancement: a deep learning approach / J. Chen, J. Liu, C. Liu [et al.] // *arXiv preprint arXiv:2104.12362* (2021)
3. Underwater acoustic target recognition method based on a joint neural network / X. C. Han, C. Ren, L. Wang, Y. Bai // *PLoS ONE* 2022. 17(4): e0266425.
4. Convolutional Neural Network With Second-Order Pooling for Underwater Target Classification / X. Cao, R. Togneri, X. Zhang, Y. Yu // *IEEE Sensors Journal*. 2019. Vol. 19, no. 8. P. 3058–3066.
5. A new cooperative deep learning method for underwater acoustic target recognition / Yang Honghui, Xu Guanghui, Yi Shuzhen, Li Yiqing // *OCEANS 2019-Marseille / IEEE*. 2019. P. 1–4.
6. Underwater target classification at greater depths using deep neural network with joint multiple-domain feature / Cao Xu, Zhang Xiaomin, Togneri Roberto, Yu Yang // *IET Radar, Sonar & Navigation*. 2019. Vol. 13, no. 3. P. 484–491.
7. DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification / M. Irfan, Z. Jiangbin, Sh. Ali [et al.] // *Expert Systems with Applications*, 2021. Vol. 183/ 115270.
8. An underwater vessel noise database / S.-D. David, T.-G. Soledad, C.-L. Antonio, P.-G. Antonio // *Applied Acoustics*. 2016. Vol. 113. P. 64–9.

A. O. Oleksii, A. A. Verlan

APPLICATION OF A MULTILAYER PERCEPTRON WITH ADAPTIVE LEARNING FOR THE ANALYSIS OF ACOUSTIC SIGNALS IN THE WATER ENVIRONMENT

Modern studies of acoustic, magnetic, and other fields of the marine environment are conducted under conditions that are characterized by a high rate of change in the external environment and an ever-increasing amount of information received on various channels in real time. This is especially typical for studies of signals with passive reception methods, which are characterized by a low SNR. Additionally, in shallow sea areas with a significant number of signal sources and serious interference, passive methods of signal monitoring encounter difficulties in noise measurement and the problem of low accuracy. Experience from such studies involving automatic informa-

tion processing reveals that classical theories are often not applicable. The most promising field of science that can efficiently address current challenges is machine learning. Machine learning offers various tools, such as neural networks, for the task of signal analysis. There are numerous variations in this field, and there are many possible solutions. However, not all solutions work well for specific tasks with various data. Hence, there is a need to enhance existing methods that can address recent challenges. A new method has been proposed, which is more efficient and provides better results in the field of underwater signal analysis. In this study, a multilayer perceptron with batch normalization and adaptive learning was proposed. These solutions surpassed the original method and demonstrated their ability to efficiently analyze data from underwater acoustic vessel sources. The accuracy of the proposed classifier reached up to 96% on a dataset containing 4 vessels with various SNR levels.

Keywords: machine learning; acoustic signals; underwater environment; classification; multilayer perceptron; adaptive learning.

