

Харківський національний університет радіоелектроніки

Факультет інформаційно-аналітичних технологій та менеджменту

Кафедра прикладної математики

Рівень вищої освіти другий (магістерський)

Спеціальність 113 Прикладна математика

(код і повна назва)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Прикладна математика

(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри ПМ _____

(підпис)

“06” листопада 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Ювченко Катерині Сергіївні

(прізвище, ім'я, по батькові)

1. Тема роботи Математичні моделі та методи розпізнавання емоцій людини на основі відеопотоку

затверджена наказом по університету від 2 листопада 2023 р. № 1276 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 10 січня 2024 р.

3. Вихідні дані до роботи відеоряд з зображенням обличчя людини

4. Перелік питань, що потрібно опрацювати в роботі _____

1. Аналіз предметної області

2. Вибір і обґрунтування методу розв'язання

3. Програмна реалізація

4. Результати обчислювального експерименту

5. Аналіз можливих застосувань

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій _____

1. Актуальність теми роботи _____

2. Постановка задачі _____

3. Аналіз предметної області _____

4. Метод чисельного аналізу _____

5. Результати обчислювального експерименту _____

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Підбір та вивчення технічної літератури за темою роботи	6 – 12 листопада 2023 р.	виконано
2	Вибір та обґрунтування методу	13 – 26 листопада 2023 р.	виконано
3	Розробка алгоритму і програми	27 листопада – 10 грудня 2023 р.	виконано
4	Проведення аналітичних досліджень та розрахунків	11 грудня – 24 грудня 2023 р.	виконано
5	Робота над текстом пояснювальної записки	25 грудня 2023 р. – 9 січня 2024 р.	виконано
6	Представлення роботи на рецензію в ЕК	10 січня 2024 р.	виконано

Дата видачі завдання 6 листопада 2023 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Єсілевський В.С.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 92 с., 2 табл., 9 рис., 1 дод., 39 джерел.

ВІДЕОПОТІК, ЕМОЦІЇ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, КОМП'ЮТЕРНИЙ ЗІР, МАШИННЕ НАВЧАННЯ, РЕГРЕСІЯ, РЕКУРЕНТНІ НЕЙРОННІ МЕРЕЖІ, РОЗПІЗНАВАННЯ ЕМОЦІЙ, ШТУЧНИЙ ІНТЕЛЕКТ.

Об'єкт дослідження – розпізнавання емоцій людини.

Мета роботи – побудова та удосконалення найбільш ефективної моделі для точного розпізнавання емоцій на основі аналізу відеофрагмента обличчя людини.

Методи дослідження – метод дослідження програмної моделі на основі згорткових та рекурентних нейронних мереж.

У даній роботі розглянуто проблему розпізнавання емоцій у відеопотоці, фокусуючись на важливих аспектах аналізу та ідентифікації емоційних станів осіб. В рамках дослідження вивчено та проаналізовано різні емоційні вирази, зокрема радість, гнів, сум, страх та інші, які можуть виявлятися у відеопотоці через обличчя та міміку.

Проблема розпізнавання емоцій у відеопотоці виникає внаслідок складнощів у розпізнанні та інтерпретації комплексних виразів обличчя та міміки людей. Ця задача важлива, оскільки емоції відіграють ключову роль у взаємодії людей, і їх розпізнавання має широкий застосунок в різних областях, від розваг та медицини до психології та інтелектуальних систем.

Однією з проблем є варіабельність виразів обличчя та відмінності в їхньому виявленні в залежності від культурних, індивідуальних та етнічних особливостей. Різноманіття факторів, таких як освітлення, кут огляду та фон, може впливати на точність розпізнавання.

Отримані результати свідчать про успішність використання комбінації згорткової та рекурентної нейронних мереж для аналізу емоційної

толерантності та збудженості на зображеннях. Модель має великий потенціал застосування в різних галузях, включаючи безпеку, медицину та маркетинг, демонструючи високу точність та адаптивність до реальних умов.

ABSTRACT

Introductory note: 92 pages, 2 tables, 9 figures, 1 appendix, 39 sources.

ARTIFICIAL INTELLIGENCE, COMPUTER VISION, CONVOLUTIONAL NEURAL NETWORKS, EMOTIONS, EMOTION RECOGNITION, MACHINE LEARNING, RECURRENT NEURAL NETWORKS, REGRESSION, VIDEO STREAM.

Object of research – human emotion recognition.

Purpose of work – building and improving the most effective model for accurate emotion recognition based on the analysis of a video fragment of a person's face.

Methods of research – a method for studying a software model based on convolutional and recurrent neural networks.

This paper considers the problem of recognizing emotions in a video stream, focusing on important aspects of analyzing and identifying emotional states of individuals. The study examines and analyzes various emotional expressions, including joy, anger, sadness, fear, and others that can be detected in a video stream through faces and facial expressions.

The problem of recognizing emotions in a video stream arises due to the difficulty in recognizing and interpreting complex facial expressions and facial expressions of people. This task is important because emotions play a key role in human interaction, and their recognition has wide applications in various fields, from entertainment and medicine to psychology and intelligent systems.

One of the challenges is the variability of facial expressions and the differences in their detection depending on cultural, individual, and ethnic characteristics. A variety of factors, such as lighting, viewing angle, and background, can affect recognition accuracy.

The results obtained indicate the success of using a combination of convolu-

tional and recurrent neural networks to analyze emotional tolerance and arousal in images. The model has great potential for application in various fields, including security, medicine, and marketing, demonstrating high accuracy and adaptability to real-world conditions.

ЗМІСТ

	С.
Перелік скорочень, умовних познач, одиниць і термінів	10
Вступ	11
1 Аналіз предметної області та постановка задач дослідження	13
1.1 Аналіз проблеми.....	13
1.2 Огляд супутніх робіт.....	17
1.3 Змістовна та формальна постановка задачі	23
1.3.1 Змістовна постановка задачі	23
1.3.2 Формальна постановка задачі	24
1.4 Постановка задач дослідження	25
2 Вибір та обґрунтування методу розв’язання	27
2.1 Основні відомості з теорії нейронних мереж в задачах розпізнавання.....	27
2.2 Типи нейронних мереж.....	29
2.2.1 Повнозв’язні нейронні мережі	29
2.2.2 Згорткові нейронні мережі	30
2.2.3 Рекурентні нейронні мережі	33
2.3 Методи попередньої обробки набору даних	35
2.4 Метрики оцінки якості регресійної моделі.....	41
Висновки за розділом 2	51
3 Програмна реалізація	52
3.1 Високорівнева мова програмування Python	52
3.2 Алгоритм розв’язання задачі розпізнавання емоцій	52
3.3 Опис програми	56
Висновки за розділом 3	58
4 Результати обчислювального експерименту та їх аналіз	60
4.1 Навчання та оцінка якості роботи моделей CNN та RNN	60
Висновки за розділом 4	61
Висновки	63

Перелік джерел посилання	64
Додаток А Лістинг програми	69

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАК, ОДИНИЦЬ І ТЕРМІНІВ

- CNN – згорткові нейронні мережі;
- DL – глибоке навчання;
- DNN – глибокі нейронні мережі;
- FER – розпізнавання виразу обличчя;
- FR – розпізнавання обличчя;
- GRU – керовані рекурентні одиниці;
- LSTM – довготривала короткочасна пам’ять;
- MAE – середня абсолютна похибка;
- MLP – повнозв’язні нейронні мережі;
- ML – машинне навчання;
- MSE – середня квадратична похибка;
- RMSE – середньоквадратична похибка;
- RNN – рекурентні нейронні мережі;
- STN – трансформерні нейронні мережі;
- ViT – трансформер зору.

ВСТУП

Актуальність теми. Зі зростанням кількості магазинів та публічних просторів, спостереження за емоційним станом людей стає важливою задачею.

З прогресом у розробці математичних моделей та методів розпізнавання емоцій на основі відеопотоку виникає можливість автоматизації цього процесу. Застосування штучного інтелекту дозволяє аналізувати величезні обсяги відеоданих у режимі реального часу, що перевищує можливості людського спостереження.

Наприклад, в системах безпеки та публічних місцях використання автоматизованого розпізнавання емоцій може значно поліпшити спроможність вчасно виявляти ситуації, пов'язані із злочинністю чи загрозою. Це дозволяє ефективно використовувати ресурси та забезпечує швидше реагування на емоційні прояви людей, допомагаючи в запобіганні негативних подій.

Розпізнавання емоцій грає важливу роль для психологічних та медичних досліджень. Це може сприяти розумінню психічного здоров'я, ефективності терапій та вивченню динаміки емоційних реакцій на різноманітні стимули.

Технології розпізнавання емоцій також можуть бути використані в системах безпеки для виявлення підозрілої чи загрозової поведінки. Аналіз емоцій у публічних місцях може слугувати інструментом відслідковування настрою групи людей та може бути корисним для забезпечення безпеки.

Також, у сфері інтернет-маркетингу та реклами використання розпізнавання емоцій дозволяє оцінювати реакцію користувачів на рекламні матеріали, що може покращити ефективність рекламних кампаній та підвищити взаємодію з аудиторією.

Загалом, розпізнавання емоцій на основі відеопотоку відкриває нові можливості у багатьох галузях, надаючи інструменти для покращення технологічних, психологічних та медичних аспектів взаємодії людини з оточуючим середовищем.

Мета і завдання кваліфікаційної роботи. Метою кваліфікаційної роботи є побудова та удосконалення найбільш ефективної моделі для точного розпізнавання емоцій на основі аналізу відеофрагмента обличчя людини.

Для досягнення поставленої мети необхідно виконати наступні завдання:

- провести огляд і аналіз сучасного стану задачі «розпізнавання емоцій»;
- розглянути існуючі методи для вирішення задачі;
- побудувати модель глибокого навчання;
- розробити програмну реалізацію побудованої моделі;
- навчити модель глибокого навчання з підбором гіперпараметрів для найбільшої ефективності;
- зробити оцінку якості роботи моделі на реальних даних;
- на основі отриманих даних зробити висновок про проведену роботу.

Об'єктом дослідження є розпізнавання емоцій людини.

Предметом дослідження є програмна реалізація моделі глибокого навчання.

Методи дослідження. У кваліфікаційній роботі використовується метод дослідження програмної моделі на основі згорткових та рекурентних нейронних мереж.

Публікації. Результати, отримані у кваліфікаційній роботі, було опубліковано у журналі «Innovative technologies and scientific solutions for industries» [1] та на міжнародних науково-технічних конференціях «Інформаційні технології та комп'ютерне моделювання» [2], «Information Systems and Technology (IST-2022)» [3] та на 27-му Міжнародному молодіжному форумі «Радіоелектроніка та молодь у XXI столітті» (м. Харків, 10-12 травня 2023 р.) [4]. Також результати роботи були подані на конкурс студентських наукових робіт зі штучного інтелекту та посіла друге місце.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Аналіз проблеми

Проблеми розпізнавання завжди були проблемами комп'ютерного зору та розпізнавання образів. Зокрема, розпізнавання обличчя або виразу обличчя вважається найбільш широко дослідженою темою в наукових і промислових сферах. Комп'ютерний зір, розпізнавання образів і технології, пов'язані з зображеннями, досягли вражаючої продуктивності з кількісної та якісної точок зору за останні роки з появою наскрізних систем навчання, таких як глибокі моделі нейронних мереж. Серед численних практичних застосувань комп'ютерного зору та розпізнавання образів автоматична та інтелектуальна міміка на основі зображень вважається однією з найпопулярніших тем, оскільки вираз обличчя передає емоційні стани та може відігравати ключову роль у виявленні, аналізі та прогнозуванні емоційних станів. У розпізнаванні виразу обличчя (англ. Facial Expression Recognition) дослідники зазвичай мали справу з окремими емоціями обличчя, такими як щастя, здивування, нейтральність, смуток, страх, огида та гнів [5]. Таким чином, FER має на меті досягти точної класифікації серед різних виразів обличчя, тобто максимізувати дистанцію між класами та мінімізувати дистанцію всередині класу.

За кілька десятиліть було запропоновано численні підходи, які поділяються на дві групи: звичайні та підходи, засновані на глибокому навчанні. Подібно до сфери розпізнавання об'єктів, звичайний FER зазвичай складається з трьох основних етапів:

а) попередня обробка зображення, що містить зображення обличчя, з подальшим виявленням області обличчя;

б) вилучення рис обличчя;

в) класифікація та розпізнавання виразів. З технологічної точки зору FER подібний до розпізнавання обличчя (англ. Facial Recognition), але FER

відрізняється від FR тим, що FER в основному має справу з сімома цільовими емоціями, згаданими вище. Крім того, міміка відіграє більш важливу роль у людському спілкуванні чи інших взаємодіях між людиною й машиною, та людиною з людиною. FR зазвичай відіграє ключову роль в ідентифікації або автентифікації людини, а не в діяльності взаємодії [6].

Системи відеоспостереження зазвичай використовуються в різних місцях для забезпечення безпеки, особливо в місцях з великими групами людей, таких як громадські місця, транспортні системи та робочі місця. Однак ці системи стикаються з декількома проблемами, включаючи необхідність моніторингу та ефективного реагування на потенційні загрози в реальному часі. Однією з головних проблем традиційних систем відеоспостереження є те, що вони покладаються на людей-операторів для моніторингу та інтерпретації поведінки людей у реальному часі. Такий результат може бути складним, особливо в багатолюдному середовищі, де буває важко відрізнити нормальну поведінку від підозрілої. Крім того, оператори можуть страждати від втоми або неувважності, що призводить до пропуску або затримки відповідей.

Технології розпізнавання емоцій можуть допомогти вирішити вищезазначені проблеми шляхом автоматичного аналізу виразу обличчя [7] і мови тіла людей у режимі реального часу для виявлення потенційних емоційних станів, які можуть вказувати на загрозу чи підозрілу поведінку. Такі технології використовують алгоритми машинного навчання (ML) для виявлення змін у виразі обличчя, таких як похмурість, підняті брови або примружені очі, що може вказувати на негативні емоції, такі як гнів, страх або смуток. Використовуючи технологію розпізнавання емоцій у системах відеоспостереження, співробітники служб безпеки можуть отримувати сповіщення або сповіщення, коли виявляється потенційно підозріла поведінка. Це може забезпечити більш цілеспрямовану реакцію та підвищити ефективність заходів безпеки. Наприклад, співробітники служб безпеки можуть бути попереджені про потенційні загрози до їх ескалації, що дозволить швидше та ефективніше реагувати. Технології з підходом до розпізнавання емоцій

також можуть допомогти підвищити громадську безпеку шляхом ідентифікації осіб, які можуть потребувати додаткової підтримки чи втручання, наприклад тих, хто переживає стрес або має проблеми з психічним здоров'ям. Це може бути корисним у стресових середовищах, таких як аеропорти, вокзали або багатолюдні громадські заходи, де люди можуть бути більш схильні до емоційного стресу [8].

Раніше для деяких досліджень використовували методи розпізнавання емоцій на основі ML у системах відеоспостереження для великих груп людей. Наприклад, дослідження [9] вивчало використання технології розпізнавання виразу обличчя для виявлення агресії та насильства в людних громадських місцях. Система використовувала традиційний алгоритм ML для аналізу виразу обличчя та виявлення потенційних загроз, таких як злість або агресивна поведінка. Система була протестована в симульованому середовищі та досягла рівня точності понад 90%. В іншому дослідженні [10] було розроблено систему розпізнавання емоцій у реальному часі, яка могла б аналізувати емоційний стан людей у багатолюдному середовищі. Система використовувала комбінацію розпізнавання виразу обличчя та аналізу мови тіла для виявлення потенційних загроз або емоційного стресу. Він був протестований у торговому центрі та досяг рівня точності понад 80%. Інша робота досліджувала використання технології розпізнавання емоцій у системах відеоспостереження для забезпечення безпеки на робочому місці. Система використовувала розпізнавання виразу обличчя та аналіз фізіологічного сигналу для виявлення потенційних загроз безпеці та сповіщення працівників у режимі реального часу. Система була випробувана на виробничому підприємстві та досягла рівня точності понад 90%.

З іншого боку, останні дослідження дослідили потенціал впровадження технології розпізнавання емоцій на основі глибокого навчання (англ. Deep Learning) у системах відеоспостереження для великих груп людей. Ці дослідження показали багатообіцяючі результати у виявленні таких емоцій, як гнів, страх, радість і смуток за виразом обличчя. Наприклад, дослідження,

опубліковане в [11], вивчало використання алгоритмів розпізнавання емоцій на основі DL для виявлення агресії та насильства в людних громадських місцях. Система змогла виявити агресію з точністю понад 90%, аналізуючи вираз обличчя, отриманий безпілотними літальними апаратами. Подібним чином інша робота продемонструвала потенціал розпізнавання емоцій на основі DL (англ. Deep Learning) у виявленні депресії та тривоги серед студентів коледжу за допомогою міміки. Система досягла рівня точності понад 80% у визначенні депресії та тривоги за виразом обличчя.

У вищезазначеному відношенні використання високоточних екстракторів орієнтирів обличчя все більше стає центральним аспектом виділення геометричних ознак у розпізнаванні емоцій. Наприклад, Каземі та Салліван [12] запропонували метод досягнення точного вирівнювання обличчя всього за одну мілісекунду. Цей підхід використовує ансамбль регресійних дерев для прогнозування позицій орієнтирів обличчя, пропонуючи значну перевагу у швидкості та точності. Таке швидке, точне вирівнювання обличчя може бути корисним для розпізнавання емоцій у реальному часі, оскільки воно сприяє швидкому виділенню геометричних характеристик, які безпосередньо сприяють розпізнаванню емоційного стану. Ефективно фіксуючи часову динаміку, RNN можуть моделювати довготривалу емоційну поведінку на основі послідовності орієнтирів обличчя. Цей підхід дозволяє точніше розпізнавати емоції з нюансами, враховуючи часові варіації виразу обличчя, покращуючи розуміння емоційних станів з часом.

Тим не менш, незважаючи на багатообіцяючі результати, згадані вище, таким дослідженням та іншим пов'язаним дослідженням часто бракує інтерпретації та прозорості їх висновків. Може бути проблематично пояснити, як сучасні моделі DL прийшли до своїх висновків, що викликає занепокоєння щодо потенційних упереджень та етичних міркувань, викликаних новою концепцією FACTS (англ. Fairness, Accountability, Confidentiality, Transparency, and Safety) [13] в ШІ. Крім того, DL-моделі вимагають великої кількості даних для навчання, що ускладнює забезпечення того, щоб моделі точно фіксували

різноманітність емоцій і виразів. Це може призвести до потенційних упереджень і обмежень у здатності цих моделей узагальнювати різні ситуації в громадських місцях. Таким чином, існує потреба в продовженні досліджень розробки інтерпретованих і більш прозорих моделей розпізнавання емоцій, які можуть ефективно виявляти емоції в системах відеоспостереження безпеки.

Проведений аналіз виявив та висвітлив наступні візуальні риси обличчя, які можуть свідчити про аномальну поведінку групи людей, які необхідно якнайшвидше виявити з відеопотоку системи відеоспостереження:

- швидка зміна емоційного стану: у звичайній розмові або взаємодії люди зазвичай не швидко переходять від однієї емоції до іншої, тобто якщо система виявляє, що вираз обличчя людини швидко змінюється між щастям, гнівом, страхом тощо, це може означати ненормальну поведінку;

- висока інтенсивність міміки: у громадських місцях люди зазвичай демонструють помірний рівень емоцій, наприклад якщо один або група людей часто виявляють надмірні емоції, наприклад дуже високий рівень страху, це може вказувати на ненормальну поведінку, спричинену надзвичайною ситуацією;

- неузгодженість міміки: якщо різні частини обличчя виражають різні або суперечливі емоції (наприклад, усмішка ротом, але не очима), це часто сприймається як ознака нещирої або ненормальної емоційної поведінки.

1.2 Огляд супутніх робіт

Для багатовимірних емоцій у просторі валентно-збудження хороший спосіб передати вхідні зображення в моделі глибокого навчання – перетворити їх у послідовності, подібні до звичайного порядку кадрів у відео. Навчання на таких даних може використовувати тимчасову динаміку відео [14].

Передача послідовностей нейронними мережами може здійснюватися різними способами, описаними нижче:

а) згорткові та рекурентні нейронні мережі пройшли спільне навчання – Колліас та інші розробили архітектуру, використовуючи існуючі сучасні архітектури CNN разом з різними RNN для завдання Aff-Wild2. Вкладена модель CNN попередньо навчається на різних наборах даних із особами або випадковими зображеннями. Потім CNN та RNN спільно навчаються у певній послідовності для досягнення хороших результатів на наборах даних RECOLA [15] та Aff-Wild2 [16], [17].

Цей спосіб навчання має такі недоліки, як складність загальної комбінованої моделі та, отже, дорожня обчислювальних ресурсів. Крім того, вони використовували механізм уваги, накладаючи шар уваги поверх RNN. Механізм уваги вирішує проблему обмеженості короткочасної пам'яті RNN. Вищезгаданий механізм є інновацією, що змінює правила гри та вирішує цю проблему;

б) згорткові та рекурентні нейронні мережі пройшли навчання окремо – Цей спосіб навчання має такі недоліки, як складність загальної комбінованої моделі та, отже, дорожня обчислень. Інший підхід представлений в [18], де автор адаптує модель CNN для роботи з даними про валентність та збудження з набору даних RECOLA. Останній щільний шар був змінений за допомогою активації softmax для категоріальних емоцій та шару регресії для роботи з просторовими емоціями. Вони запропонували механізм увімкнення тимчасової інформації за допомогою RNN для поширення інформації з одного моменту часу в інший. Кожен вхід у RNN складається з функцій з другого до останнього повністю зв'язного рівня одного кадру CNN. Основна відмінність цього підходу від попередніх полягає у роздільному навчанні CNN та RNN. Вони оцінюють, який внесок CNN і RNN окремо вносять у загальну продуктивність. Вони розширюють цей підхід, беручи до уваги аудіо та фізіологічні дані. Він виявив корисність різних функцій для прогнозування та досліджував, як додавання аудіо та фізіологічних функцій впливає на продуктивність. Крім того, вони використовували механізм уваги, накладаючи шар уваги поверх RNN. Механізм уваги вирішує проблему обмеженості короткочасної пам'яті RNN.

Вищезгаданий механізм є інновацією, що змінює правила гри та вирішує цю проблему.

Архітектура трансформера, яка широко використовується в області обробки природної мови (англ. Natural language processing), демонструє винятково хороші результати в задачі розпізнавання, особливо у випадку використання послідовностей зображень як вхідних даних [19]. Нещодавно був запропонований новий навчальний модуль, мережа просторових трансформерів (англ. Spatial transform networks), щоб забезпечити надійну продуктивність, дозволяючи просторове маніпулювання вхідними даними зображення. STN вбудовується в існуючі мережеві моделі, наприклад, CNN, і це дозволяє досягти надійних результатів навчання від інваріантності до просторової трансформації вхідних даних, наприклад, трансляції, повороту, деформації тощо. Модель STN була застосована до багатьох практичних проблем серед багатьох випадків проблем розпізнавання з обнадійливими результатами.

Інша модель трансформера, трансформер зору (ViT), привернула увагу в галузі розпізнавання та була запропонована як альтернатива існуючим моделям DNN.

Незважаючи на те, що існуюча робота FER досягла значно кращих результатів з кількісної та якісної точок зору, це все ще є складним завданням через існування неконтрольованого зовнішнього середовища, варіацій поз або оклюзії, які погіршують ефективність результатів FER. Більш складні сценарії FER потрібно розглядати для отримання високоякісних FER з практичної точки зору. Таким чином, варто дослідити методи FER з використанням STN, який привернув увагу в області глибокого навчання. У цьому документі, натхненному мережевим модулем просторового трансформера, FER виконується ефективно шляхом вибору найбільш релевантної частини зображення обличчя з подальшим застосуванням триплетної функції втрат. Результат порівнюється з результатами з використанням крос-ентропії та з іншими результатами FER з використанням найсучасніших алгоритмів. Запропонований метод показує чудові результати розпізнавання в FER,

особливо у випадку зображень обличчя, які мають зони оклюзії. Мережа просторового трансформера, що входить до стандартної структури нейронної мережі, має переваги у разі обертання, обрізання, масштабування та нежорсткої деформації зображень, які на практиці інколи трапляються навпроти.

Як добре відомо, традиційні методи FER використовують підходи на рівні пікселів, геометричні моделі або підходи на рівні об'єктів. Сучасні підходи зазвичай класифікують як підходи на основі глибокого навчання. За останні кілька років підходи, засновані на навчанні, свідчать про значне вдосконалення завдань розпізнавання, особливо в сферах розпізнавання обличчя, розпізнавання виразу обличчя, розпізнавання діяльності тощо. У підходах, заснованих на глибокому навчанні, згорткові нейронні мережі та повторювані моделі нейронних мереж є найбільш широко використовуваними. Останнім часом трансформерна мережа вважалася однією з альтернатив підходам на основі CNN і RNN. Модель глибокої нейронної мережі дозволяє виконувати автоматизоване FER, що вже давно є цікавим і складним завданням у сфері проблем розпізнавання. Замість вилучення ознак із зображень обличчя за допомогою певної математичної чи статистичної моделі підхід до розпізнавання на основі DNN виділяє різноманітні та численні ознаки за допомогою великої кількості прихованих шарів, які сприяють виділенню ознак за допомогою механізму, подібного до мозку.

На відміну від традиційного підходу, який завжди намагається досягти мінімальної дистанції між класами та максимальної дистанції між класами в проблемі розпізнавання за допомогою аналітичної моделі (математики, статистики тощо), останній підхід, заснований на глибокому навчанні, здатний знайти абстракцію. і складні візерунки, які притаманні реальним зображенням обличчя.

Традиційним підходом до розпізнавання обличчя, виразу обличчя, діяльності чи об'єкта не вистачає можливості узагальнення через варіації пози чи масштабу та довільний додатковий шум. Крім того, через відсутність наборів даних майже всі завдання розпізнавання ґрунтувалися на ручному або

аналітичному виділенні точок ознак на основі моделі. Натхненні появою глибокого навчання, моделі на основі CNN показали стійкість до вищезгаданих варіацій, тому FER застосував модель CNN для досягнення вищої точності рівня розпізнавання. Аналіз даних обличчя на основі CNN з'явився в роботі Лоуренса, Лекуна і Фазеля, робота яких використовувала менше п'яти прихованих шарів у їхніх мережевих моделях. Майже всі алгоритми FER також використовують ці роботи як основу, щоб запропонувати новизну або подальше покращення точності розпізнавання. З моменту роботи FER на ранніх етапах було досягнуто значного прогресу в більш практичних і диких обставинах завдяки використанню моделі DNN. Модель CNN є однією з найперших, яка глибоко вивчає та виділяє риси обличчя з тонкими змінами виразу, які важко виділити за допомогою традиційних методів розпізнавання. З моменту впровадження CNN для роботи з розпізнавання FER також використовував структуру CNN, додавши більше рівнів, що призвело до глибоких архітектур CNN, які покращили результати FER. На початковій стадії FER з використанням CNN використовувалась обмежена кількість наборів даних зображень. Було запропоновано, що незалежність суб'єкта та трансляція, обертання та масштабно-інваріантний FER з використанням CNN відрізняють посмішку від розмови на основі помітності візуальних ознак [20]. Натхненний вираженням справжніх емоцій, FER був розширений до розпізнавання мікровиразів за допомогою методів глибокого навчання. Єдина мережева структура глибокого навчання, яка складається з двох шарів згортки з наступним максимальним об'єднанням і чотирьох початкових рівнів, була представлена на ранніх етапах FER за допомогою DNN, але ця робота використовує зареєстрований набір даних зображення обличчя, а орієнтир витягується ап'юрі. Було проведено багато досліджень для вирішення проблем FER дикого набору даних. Робота в [21] пропонує структуру багатозадачного навчання, яка використовує залежності між цими двома моделями за допомогою графової згорткової мережі для розпізнавання виразів обличчя в дикій природі. Робота в [22] пропонує візуальну структуру наскрізного

розпізнавання емоцій, яка складається з надійної попередньо навченої базової моделі та тимчасової підсистеми для моделювання часових залежностей між багатьма відеокадрами. Крім того, міміку можна застосовувати в багатьох додатках. У [23] механізм уваги використовує напівконтрольований локалізатор, який точно виявляє помітні області, а модель STN вбудовується в існуючу модель DNN (наприклад, модель CNN) для вирішення проблеми розпізнавання та виявлення у випадку просторово перетвореного вхідного сигналу. зображення. Робота показує, що STN входить до базової моделі CNN, і вся архітектура складається з трьох частин: локалізації, сітки вибірки та вибірки зображення. Однак він може локалізувати приблизну позицію цілі (наприклад, номер футболки футболіста), і немає роботи над вибором адаптивної функції втрат оптимізації. Згортка уваги була введена для класифікації виразів обличчя, де кількість класів менша, ніж у звичайних випадках проблем класифікації [24]. У роботі [25] автори використали менше 10 прихованих шарів і додали механізм уваги для ефективного FER, а запропонований підхід повідомив про кращу точність, ніж результати найсучаснішого рівня. Однак точність показує коливання, і в роботі повідомляється, що існує компроміс між швидкістю розпізнавання та швидкістю конвергенції. Оклюзія або варіації пози є двома основними факторами, які погіршують точність розпізнавання, тому для надійного FER використовуються мережі концентрації уваги регіону шляхом адаптивного захоплення важливих областей обличчя. Оскільки FER у природі є складним завданням, було запропоновано механізм уваги з базовою моделлю CNN для сприйняття областей оклюзії, зосереджуючись на найбільш незакритих областях обличчя. У [26] пропонується розширення базової моделі STN шляхом додавання процедур для захоплення ефективних областей уваги за допомогою орієнтирів обличчя або карт візуальної помітності обличчя. У цій роботі FER на основі STN було додано до моделі CNN із просторовою та каналною увагою, і подальшого вдосконалення, можливо, можна досягти за допомогою запропонованої функції активації GELU (лінійна одиниця помилки Гауса). Було

запропоновано мультимодальне розпізнавання емоцій, яке використовує мову та зображення обличчя. У цьому підході для розпізнавання емоцій пропонується попередньо навчений STN для карт помітності та bi-LSTM для механізму уваги. Однак у цій роботі вхідне зображення перетворюється на мел-спектограми, що призводить до збільшення обчислювальної складності.

Незважаючи на зусилля в FER з використанням DNN з механізмом уваги, є місце для подальшого вдосконалення, і наш запропонований метод дає FER, який є стійким до оклюзії та ефективним, зосереджуючись на найбільш релевантній області обличчя для певного виразу шляхом додавання STN. На додаток до методів, що використовують STN, наш підхід використовує адаптивну функцію втрат і триплетну функцію втрат, які покращують точність розпізнавання у випадку оклюзії.

1.3 Змістовна та формальна постановка задачі

1.3.1 Змістовна постановка задачі

Проблема розпізнавання емоцій у відеопотоці визначається рядом труднощів, які виникають при спробі автоматично аналізувати та класифікувати емоційні стани осіб на відеофрагментах. Основні аспекти цієї проблеми включають варіабельність виразів обличчя, яка може значно варіюватися в залежності від індивідуальних особливостей, культурних відмінностей та інших факторів. Проблема полягає також у високій незбалансованості даних, де негативні емоції або слабо виражені вирази є рідкісними порівняно з вираженими позитивними емоціями.

Вплив факторів оточення, таких як освітлення, об'єкти на задньому плані та артефакти відеозапису, може суттєво впливати на якість та точність розпізнавання емоцій. Неоднозначність інтерпретації мимічних виразів додає до проблеми, оскільки одні й ті ж самі вирази обличчя можуть інтерпретуватися різними спостерігачами по-різному, враховуючи суб'єктивне сприйняття.

Протягом багатьох років задача розпізнавання емоцій залишалася однією з найцікавіших та важливих проблем у галузі взаємодії людина-комп'ютер. В даній кваліфікаційній роботі було розглянуто задачу розпізнавання емоцій як класифікаційне, а також як завдання регресії, обробляючи закодовані емоції в різних наборах даних за допомогою моделей глибокого навчання. Дана модель поєднує згорткову нейронну мережу (CNN) з рекурентною нейронною мережею (RNN) для прогнозування емоцій у відеоданих. На першому етапі CNN витягує вектори ознак з кадрів відео. На другому етапі ми передаємо ці вектори ознак для навчання RNN з метою використання часової динаміки відео. Крім того, було проаналізовано, як кожна нейронна мережа сприяє загальній продуктивності системи. Експерименти виконані на публічно доступних наборах даних, включаючи найбільш сучасну базу даних Aff-Wild2. Вона містить понад шістдесят годин відеоданих. Було виявлено проблему перенавчання моделі на незбалансованому наборі даних з ілюстративним прикладом, використовуючи матриці невідповідностей. Проблема вирішена за допомогою техніки зменшення вибірки для забезпечення балансу набору даних. Значно зменшивши тренувальні дані, було збалансовано набір даних, тим самим покращивши загальну продуктивність моделі. Отже, дослідження якісно описує можливості моделей глибокого навчання для вивчення достатньої кількості даних для прогнозування емоцій обличчя.

1.3.2 Формальна постановка задачі

Формальна постановка задачі CNN-RNN для залежних змінних може бути представлена в наступному вигляді.

Вхідні дані – набір зображень.

Вихідні дані – значення двох залежних змінних y_1 і y_2 для кожного зображення в наборі X .

Нехай $x \in X$ представляє собою зображення у форматі $f \times h \times w \times c$, де h , w і c – висота, ширина і кількість каналів зображення відповідно. Тоді функція $f(x)$ може бути наступним чином:

$$f(x) = (y_1(x), y_2(x)), \quad (1.1)$$

де $y_1(x)$ і $y_2(x)$ – значення залежних змінних для зображення x .

CNN в нашому випадку використовується для вилучення ознак із зображення x . Нехай $F(x)$ представляє собою вихід CNN для зображення x . Тоді функція $f(x)$ може бути представлена у наступному вигляді:

$$f(x) = (y_1(F(x)), y_2(F(x))), \quad (1.2)$$

де $y_1(F(x))$ і $y_2(F(x))$ – значення залежних змінних для $F(x)$.

RNN використовується для передбачення значень залежних змінних на основі отриманих ознак. Нехай h_t представляє собою стан RNN в момент часу t . Тоді функція $f(x)$ виглядає так:

$$f(x) = (y_1(h_T), y_2(h_T)), \quad (1.3)$$

де h_T – стан RNN в кінці відеоряду.

Метою є знайти функцію $f(x)$, яка відображає вхідне зображення $x \in X$ на пару значень залежних змінних y_1 і y_2 .

1.4 Постановка задач дослідження

Метою дослідження є побудова та удосконалення найбільш ефективної моделі для точного розпізнавання емоцій на основі аналізу відеофрагмента

обличчя людини. Сформулюємо задачі для дослідження в рамках даної кваліфікаційної роботи:

- провести огляд і аналіз сучасного стану задачі «розпізнавання емоцій»;
- розглянути існуючі методи для вирішення задачі;
- побудувати модель глибокого навчання;
- розробити програмну реалізацію побудованої моделі;
- навчити модель глибокого навчання з підбором гіперпараметрів для найбільшої ефективності;
- зробити оцінку якості роботи моделі на реальних даних;
- на основі отриманих даних зробити висновок про проведену роботу.

2 ВИБІР ТА ОБҐРУНТУВАННЯ МЕТОДУ РОЗВ'ЯЗАННЯ

2.1 Основні відомості з теорії нейронних мереж в задачах розпізнавання

Нещодавнє зростання інтересу до нейронних мереж пов'язане з усвідомленням того, що мозок виконує обчислення інакше, ніж звичайні цифрові комп'ютери. Комп'ютери надзвичайно швидкі і точні у виконанні послідовностей інструкцій, які були для них сформульовані. Інформаційна обробка людини складається з нейронів, які перемикаються зі швидкістю, приблизно в мільйон разів меншою, ніж ворота комп'ютера. Однак люди є більш ефективними ніж комп'ютери у вирішенні складних обчислювальних завдань, таких як розуміння мови.

Штучні нейронні мережі – це фізичні клітинні системи, які можуть отримувати, зберігати та використовувати отримані знання. Знання представлені у вигляді стабільних станів чи відображень, вбудованих у мережі, які можуть бути відновлені відповідно до пред'явлених сигналів. Обробка нейронних мереж, як правило, включає роботу з проблемами великого масштабу щодо розмірності, обсягу оброблюваної інформації та обсягу симуляції чи обробки нейронної апаратури.

Нейронна мережа – це обчислювальна система, яка складається з числа простих, взаємопов'язаних вузлів або обчислювальних елементів, які обробляють інформацію за допомогою їхнього динамічного стану відповіді на зовнішні вхідні сигнали [27].

Штучні нейронні мережі, які знаходять своє натхнення в структурі людського мозку, є ефективним інструментом для розв'язання задач машинного навчання та штучного інтелекту. Зазначена система відображає структурні та функціональні аспекти біологічного мозку, складаючись з абстрагованої моделі взаємопов'язаних нейронів. Ця модель може бути успішно використана для вирішення різноманітних завдань у різних галузях, таких як статистика, технології та економіка.

Нейронні мережі, як об'єкт досліджень у галузі нейроінформатики та штучного інтелекту, вимагають попереднього навчання перед тим, як вони зможуть ефективно розв'язувати задачу. Структура та принцип дії нейронної мережі можуть бути визначені наступним чином: абстрактна модель містить нейрони, або одиниці, які можуть збирати та передавати інформацію. Вхідні нейрони отримують інформацію від зовнішнього середовища у вигляді патернів чи сигналів. Приховані нейрони, що знаходяться між вхідними та вихідними, виконують функцію відображення внутрішніх патернів інформації. Вихідні нейрони передають інформацію до зовнішнього світу як результат роботи мережі [28].

Класифікація є однією з найбільш популярніших задач прийняття рішень у людській діяльності. Проблема класифікації виникає, коли об'єкт потрібно віднести до попередньо визначеної групи або класу на основі численних спостережуваних атрибутів, пов'язаних з цим об'єктом. Багато проблем у бізнесі, науці, промисловості та медицині можуть бути розглянуті як проблеми класифікації. Приклади включають прогноз банкрутства, кредитний рейтинг, медичну діагностику, контроль якості, розпізнавання рукописного тексту та розпізнавання мови.

Традиційні статистичні процедури класифікації, такі як дискримінантний аналіз, базуються на байєсівській теорії прийняття рішень. У цих процедурах потрібно припустити підґрунтьову ймовірнісну модель, щоб розрахувати апостеріорну ймовірність, на якій ґрунтується рішення класифікації. Однак одним з основних обмежень статистичних моделей є те, що вони працюють ефективно лише при виконанні передбачених припущень. Ефективність цих методів значною мірою залежить від різних припущень чи умов, за яких розробляються моделі. Користувачам необхідно мати глибоке розуміння як властивостей даних, так і можливостей моделей, перш ніж моделі можна успішно застосовувати.

Штучні нейронні мережі визначаються як важливий інструмент для класифікації. Недавні дослідження в галузі нейронної класифікації свідчать про

те, що вони є перспективною альтернативою різноманітним традиційним методам класифікації. Перевага нейронних мереж полягає в теоретичних аспектах.

По-перше, нейронні мережі є методами, орієнтованими на дані, що можуть адаптуватися до даних без якої-небудь явної специфікації функціональної чи розподілової форми підґрунтевої моделі.

По-друге, вони є універсальними функціональними апроксиматорами, оскільки можуть наближати будь-яку функцію з довільною точністю. Оскільки будь-яка процедура класифікації прагне до функціонального відношення між членством у групі та атрибутами об'єкта, точне визначення цієї підґрунтевої функції надзвичайно важливо.

По-третє, нейронні мережі є нелінійними моделями, що робить їх гнучкими в моделюванні складних взаємозв'язків реального світу. Нарешті, нейронні мережі здатні оцінювати апостеріорні ймовірності, що надає основу для встановлення правил класифікації та виконання статистичного аналізу [28].

2.2 Типи нейронних мереж

2.2.1 Повнозв'язні нейронні мережі

Модель багатошарового персептрону (повнозв'язна нейронна мережа) є типом штучної нейронної мережі прямого поширення, яка служить базовою архітектурою для глибокого навчання або глибоких нейронних мереж. Він працює як контрольований підхід до навчання. MLP складається з трьох шарів: вхідного, вихідного та одного або декількох прихованих шарів. Це повністю зв'язна мережа, тобто кожен нейрон одного шару пов'язаний із усіма нейронами наступного шару [29].

У MLP вхідний рівень отримує вхідні дані та виконує нормалізацію ознак. Приховані шари, кількість яких може бути різною, обробляють вхідні

сигнали. Вихідний рівень приймає рішення чи прогнози з урахуванням обробленої інформації.

На рисунку 2.1 зображено модель перцептрон з одним нейроном, де функція активації φ (2.1) є нелінійною функцією, що використовується для зіставлення функції підсумовування $(xw + b)$ з вихідним значенням y :

$$y = \varphi(xw + b). \quad (2.1)$$

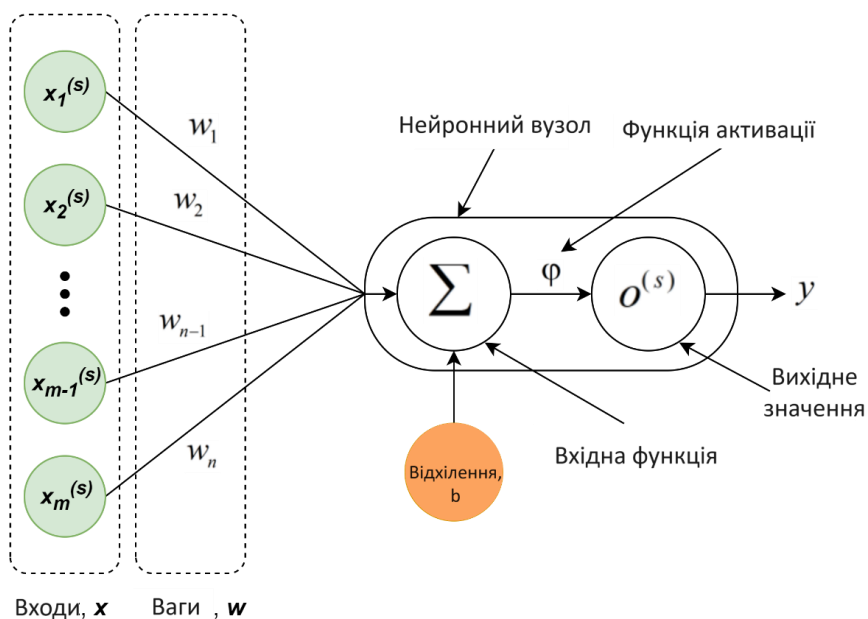


Рисунок 2.1 – Модель перцептрон з одним нейроном

У (2.1) терміни x , w , b і y є вхідний вектор, ваговий вектор, зсув і вихідне значення відповідно [30]. Рисунок 2.2 ілюструє структуру моделі багатошарового перцептрона.

2.2.2 Згорткові нейронні мережі

Згорткові нейронні мережі є потужним класом моделей глибокого навчання, що широко застосовуються в різних задачах, включаючи виявлення

об'єктів, розпізнавання мови, комп'ютерний зір, класифікацію зображень і біоінформатику. Вони також продемонстрували успіх у задачах прогнозування часових рядів. CNN – це нейронні мережі прямого поширення, які використовують згорткові структури для отримання функцій з даних [31]. На відміну від традиційних методів, CNN автоматично вивчають та розпізнають ознаки на основі даних без необхідності ручного отримання ознак людьми. Основні компоненти CNN включають згортковий рівень, рівень пулу та повнозв'язний рівень. На рисунку 2.3 представлено типову архітектуру CNN для задач класифікації зображень.

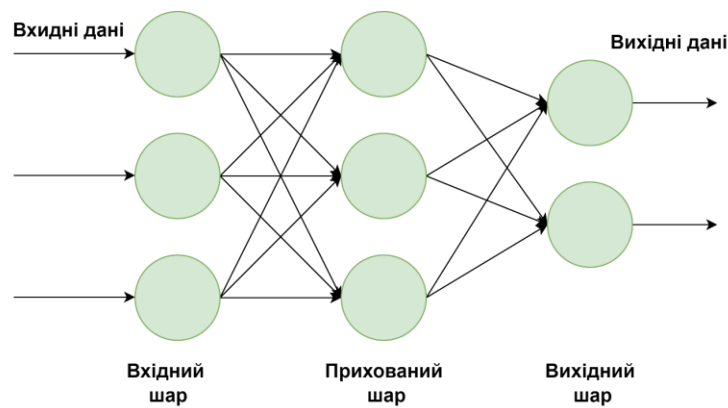


Рисунок 2.2 – Структура багатозарового персептрона

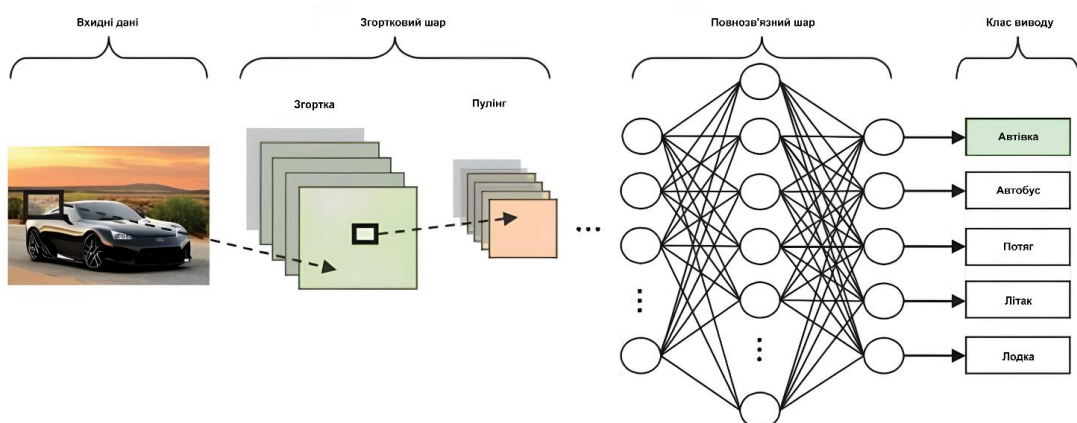


Рисунок 2.3 – Конвеєр CNN для класифікації зображень

Згортковий рівень. Згортковий шар є ключовим компонентом CNN. За допомогою кількох згорткових шарів операція згортки витягує з вхідних даних

різні функції. При класифікації зображень нижні шари, як правило, фіксують основні характеристики, такі як текстура, лінії та краї, а вищі шари отримують більш абстрактні особливості. Згортковий шар містить ядра, що навчаються, згортки, які являють собою вагові матриці, зазвичай рівної довжини, ширини і непарного числа (наприклад, 3x3, 5x5 або 7x7). Ці ядра згорнуті з вхідними картами ознак, ковзаючи по областях карти ознак і виконуючи операції згортки [32]. Рисунок 2.4 ілюструє важливу схему процесу згортки.

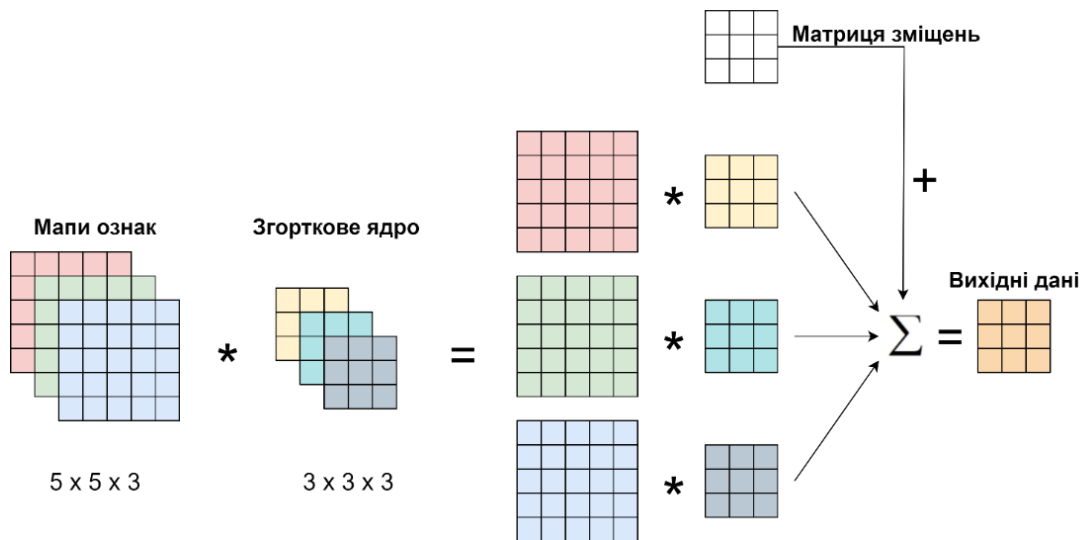


Рисунок 2.4 – Принципова схема процесу згортки

Рівень пулу. Зазвичай наступний за шаром згортки рівень пулу зменшує кількість з'єднань в мережі, виконуючи знижувальну дискретизацію і зменшення розмірності вхідних даних. Його основна мета – полегшити обчислювальне навантаження та вирішити проблеми переоснащення. Більш того, шар об'єднання дозволяє CNN розпізнавати об'єкти, навіть якщо їх форми спотворені або проглядаються під різними кутами шляхом об'єднання різних розмірів зображення за допомогою об'єднання. Операція об'єднання створює вихідні карти ознак, які стійкіші до спотворень і помилок окремих нейронів. Існують різні методи об'єднання, включаючи максимальне об'єднання, середнє об'єднання, об'єднання як просторової піраміди, змішане об'єднання, багатомасштабне безупорядковане об'єднання і стохастичне об'єднання. На

рисунку 2.5 показаний приклад Max Pooling, коли вікно ковзає за вхідними даними, а вміст вікна обробляється функцією об'єднання [33].

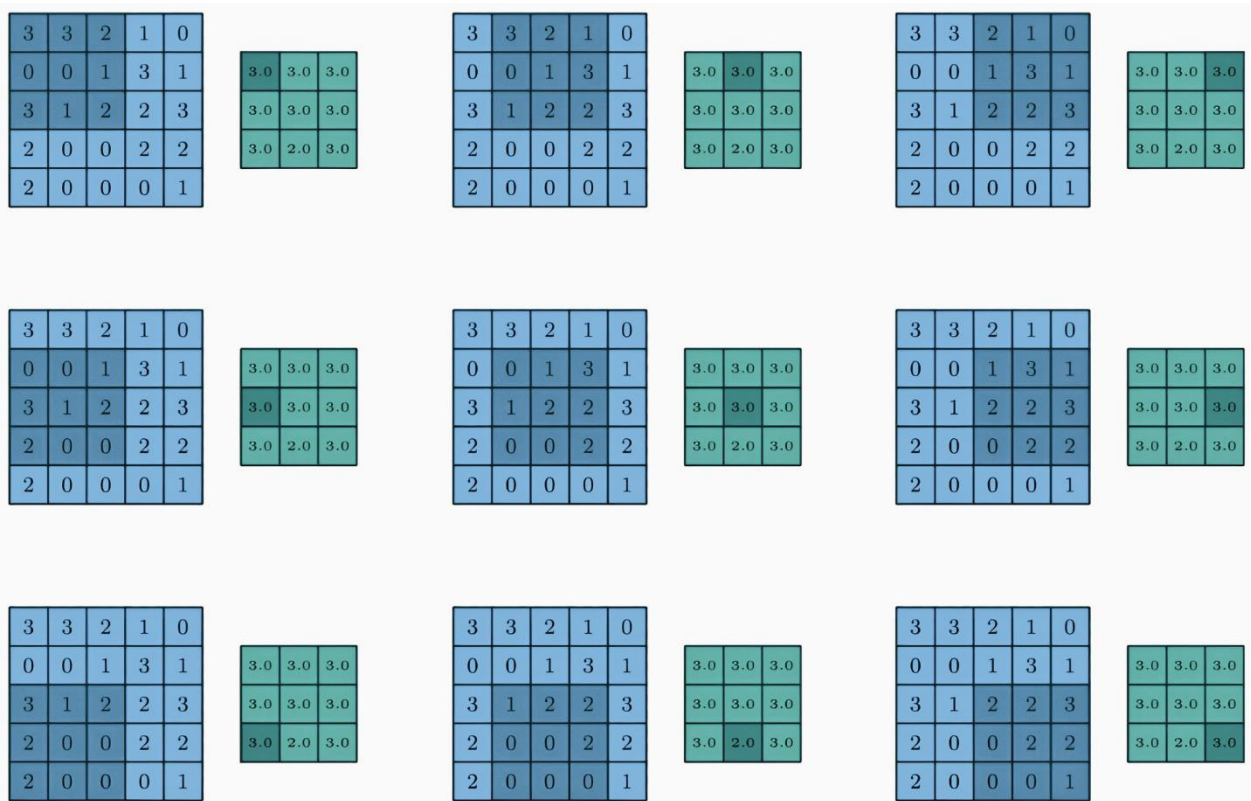


Рисунок 2.5 – Обчислення вихідних значень операції максимального об'єднання 3×3 на вході 5×5

2.2.3 Рекурентні нейронні мережі

Рекурентні нейронні мережі – це клас моделей глибокого навчання, які мають внутрішню пам'ять, що дозволяє їм вловлювати послідовні залежності. На відміну від традиційних нейронних мереж, які розглядають вхідні дані як незалежні об'єкти, RNN враховують тимчасовий порядок вхідних даних, що робить їх придатними для завдань, пов'язаних із послідовною інформацією. Використовуючи цикл, RNN застосовують ту саму операцію до кожного елемента в серії, при цьому поточне обчислення залежить як від поточного введення, так і від попередніх обчислень.

Здатність RNN використовувати контекстну інформацію особливо цінна у таких завданнях, як обробка природної мови, класифікація відео та розпізнавання мови. Наприклад, при моделюванні мови розуміння попередніх слів у реченні має вирішальне значення для передбачення наступного слова. RNN чудово вловлюють такі залежності завдяки своїй природі, що повторюється.

Однак обмеженням простих RNN є їхня короткочасна пам'ять, яка обмежує їхню здатність зберігати інформацію в довгих послідовностях. Щоб подолати цю проблему, були розроблені більш просунуті варіанти RNN, у тому числі Long ShortTerm Memory (LSTM), двонаправлений LSTM, GRU, двонаправлений GRU, байсовський RNN, та інші.

На рисунку 2.6 зображено просту рекурентну нейронну мережу, в якій внутрішня пам'ять (ht) обчислюється з використанням рівняння (2.2) [34]:

$$h_t = g(Wx_t + Uh_t + b), \quad (2.2)$$

де $g(Wx_t + Uh_t + b)$ – функція активації (зазвичай гіперболічний тангенс);

U і W – регульовані вагові матриці для прихованого стану;

h , b – член зміщення;

x – вхідний вектор.

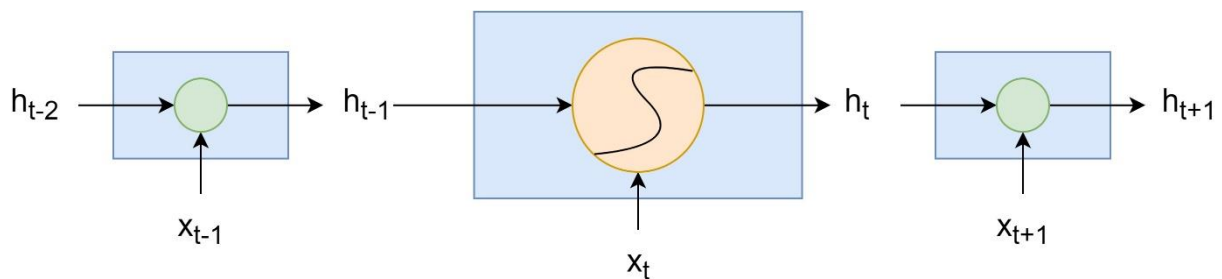


Рисунок 2.6 – Проста внутрішня робота RNN

RNN зарекомендували себе як потужні моделі обробки послідовних даних, що ефективно використовують свою здатність фіксувати залежності з

часом. Різні типи моделей RNN, такі як LSTM, двонаправлений LSTM, GRU та двонаправлений GRU, були розроблені для вирішення конкретних завдань у різних додатках [29].

2.3 Методи попередньої обробки набору даних

Практично будь-який тип аналізу даних, науки даних або розробки штучного інтелекту вимагає певного типу попередньої обробки даних, щоб забезпечити надійні, точні та надійні результати.

Реальні дані безладні та часто створюються, обробляються та зберігаються різними людьми, бізнес-процесами та додатками. В результаті в наборі даних можуть бути відсутні поля, містяться помилки ручного введення або мати повторювані дані або різні імена для опису одного і того ж об'єкта. Люди часто можуть виявити та усунути ці проблеми в даних, які вони використовують у своїй сфері бізнесу, але дані, що використовуються для навчання алгоритмів машинного навчання або глибокого навчання, повинні автоматично попередньо оброблятися, оскільки алгоритми машинного навчання та глибокого навчання працюють найкраще, коли дані представлені у форматі, який наголошує на відповідних аспектах, необхідних для вирішення проблеми [35].

Попередня обробка даних – це метод інтелектуального аналізу даних, який тягне за собою перетворення необроблених даних на зрозумілий формат. Реальні дані часто є неадекватними, непослідовними або не містять конкретних видів діяльності або тенденцій, а також містять численні неточності. Це може призвести до збирання даних низької якості і, як наслідок, створення моделей низької якості на основі цих даних. Попередня обробка даних є способом вирішення таких проблем [36].

Більшість сучасних пакетів і сервісів для обробки даних тепер включають різні бібліотеки попередньої обробки, які допомагають автоматизувати багато з

цих завдань. Застосування алгоритмів інтелектуального аналізу даних до цих зашумлених даних не дасть якісних результатів, оскільки не зможуть ефективно ідентифікувати закономірності. Тому обробка даних важлива поліпшення загальної якості даних.

Одне застереження, яке слід дотримуватися під час попередньої обробки даних: можливість повторного кодування упередженості в наборі даних. Виявлення та виправлення упередженості має вирішальне значення для програм, які допомагають приймати рішення, що впливають на людей, наприклад, для схвалення кредитів. Хоча вчені, які працюють із даними, можуть навмисно ігнорувати такі змінні, як стать, раса чи релігія, ці характеристики можуть корелювати з іншими змінними, такими як поштові індекси або відвідуваність шкіл, що призводить до упереджених результатів.

Кроки, що використовуються при попередній обробці даних, включають:

а) профілювання даних. Профілювання даних – це процес вивчення, аналізу та перевірки даних для збору статистичних даних про їхню якість. Він починається з огляду існуючих даних та їх характеристик. Вчені, що працюють з даними, визначають набори даних, які мають відношення до проблеми, інвентаризують її важливі атрибути і формують гіпотезу про функції, які можуть мати відношення до запропонованої задачі аналітики або машинного навчання. Вони також пов'язують джерела даних з відповідними бізнес-концепціями та розглядають, які бібліотеки попередньої обробки можна використовувати [35];

б) очищення даних. Очищення даних виконується для вирішення цього аспекту. Методи очищення даних спрямовані на заповнення пропущених значень, згладжування шуму при виявленні викидів та усунення неточностей даних. Нечисті дані можуть заплутати дані та модель. Тому обробка даних різними методами очищення є важливим етапом попередньої обробки даних.

Є дві основні проблеми, пов'язані з очищенням даних: відсутні значення (може статися під час збору даних або в результаті застосування правила перевірки даних, але в будь-якому випадку необхідно враховувати пропущені

значення.) та дані з шумом (безглузді дані, які машини не можуть інтерпретувати, що може бути викликано поганим збором даних, проблемами з введенням даних і т. д.).

Відсутність значень в наборі даних можна вирішити одним з наступних способів:

- видалення рядків (стовпців). Якщо весь рядок має нечислові значення, він не витягає з нього ніякого значення. Тому такі рядки (стовпці) слід негайно видалити. Або якщо відсоток рядка/стовпця в основному відсутній, скажімо, більше 65%, тоді також можна відмовитися;

- перевірка дублікатів: якщо один і той же рядок або стовпець повторюється, ви також можете видалити їх, зберігши перший екземпляр. Щоб під час запуску алгоритмів машинного навчання не пропонувати цьому конкретному об'єкту даних перевагу чи упередженість;

- оцінка відсутніх значень. Якщо немає лише невеликий відсоток значень, для заповнення прогалін можна використовувати базові методи інтерполяції. Однак найбільш типовий підхід до роботи з відсутніми даними – заповнити їх середнім, медіанним чи значенням режиму.

Покращити набір даних що містить зашумлені дані можна використовуючи один з наступних способів:

- а) метод угруповання: цей метод згладжує відсортовані дані. Дані поділяються на частини однакового розміру і процес виконується з використанням різних підходів. Кожен сегмент розглядається окремо. Всі дані в сегменті можна замінити середнім значенням або використовувати граничні значення для виконання завдання;

- б) кластеризація: у цьому методі пов'язані дані групуються кластер. Викиди можуть залишитись непоміченими або вийти за межі кластерів;

- в) регресія: підганяючи дані до функції регресії, можна згладити. Модель регресії, що використовується, може бути лінійною (тільки з однією незалежною змінною) або множинною (з безліччю незалежних змінних) (з кількома незалежними змінними) [36];

- г) скорочення обсягу даних.

Розмір набору даних у сховищі даних може бути занадто великим, щоб його можна було обробляти алгоритмами аналізу та інтелектуального аналізу даних. Одним із можливих рішень є отримання скороченого представлення набору даних, який є набагато меншим за обсягом, але дає таку ж якість аналітичних результатів.

Розберемо покрокове керівництво щодо різних стратегій зменшення даних:

1) агрегація кубів даних.

Це спосіб скорочення даних, за якого зібрані дані виражаються в зведеній формі;

2) зменшення розмірності.

Методи зменшення розмірності використовуються для вилучення ознак. Розмірність набору даних відноситься до атрибутів або окремих характеристик даних. Ця техніка спрямована на зменшення кількості зайвих функцій, які ми розглядаємо в алгоритмах машинного навчання. Зменшення розмірності можна здійснити за допомогою таких методів, як аналіз головних компонентів тощо;

3) стиснення даних.

Використовуючи технології кодування, можна значно зменшити розмір даних. Але стиснення даних може бути як із втратами, так і без втрат. Якщо вихідні дані можна отримати після реконструкції зі стиснених даних, це називається скороченням без втрат; інакше це називається зменшенням втрат;

4) дискретизація.

Дискретизація даних використовується для поділу атрибутів безперервного характеру на дані з інтервалами. Це робиться тому, що безперервні ознаки мають менший шанс на кореляцію з цільовою змінною. Таким чином, може бути важче інтерпретувати результати. Після дискретизації змінної можна інтерпретувати групи, що відповідають цілі. Наприклад, вік атрибута може бути розділений на такі групи, як нижче 18, 18-44, 44-60, вище 60;

5) зменшення чисельності.

Дані можуть бути представлені у вигляді моделі або рівняння, як регресійна модель. Це позбавить від тягаря зберігання величезних наборів даних замість моделі;

б) вибір підмножини атрибутів.

Дуже важливо бути конкретним у виборі атрибутів. Інакше це може призвести до даних великої розмірності, які важко навчити через проблеми з недостатньою або переобладнаністю. Слід враховувати лише ті атрибути, які додають більшої цінності навчанню моделі, а решту можна відкинути [37].

г) перетворення даних. Цей етап використовується для перетворення даних у формат, який можна використовувати у процесі інтелектуального аналізу. Це робиться такими способами:

1) нормалізація: робиться для масштабування значень даних у вказаному діапазоні (від -1,0 до 1,0 або від 0,0 до 1,0);

2) генерація концептуальної ієрархії: використовуючи ієрархії понять, низькорівневі чи примітивні або необроблені дані замінюються концепціями вищого рівня під час узагальнення даних. Категоріальні якості, наприклад, узагальнюються до таких понять вищого рівня, як вулиця, місто та нація. Аналогічно, числові значення атрибутів можуть бути перетворені на поняття вищого рівня, такі як вік, наприклад, молодий, середній або літній;

3) згладжування: дозволяє видалити шум даних. До таких методів належать біннінг, кластеризація та регресія;

4) агрегація. Агрегація – це процес застосування операцій підсумовування чи агрегування до даних. Наприклад, дані про щоденні продажі можна об'єднати для розрахунку місячних та річних підсумків. Агрегація функцій. Якщо функції сильно корелюють або якщо функції можна об'єднати в іншу єдину функцію, то варто це зробити. Наприклад, у наборі даних є висота та ширина об'єкта, після чого їх можна об'єднати в одну область об'єктів. Це зменшує розмірність. Ці типи функцій за своєю природою сильно корельовані, у результаті вони також зменшують мультиколінеарність [36];

г) поповнення даних. На цьому етапі дослідники даних застосовують різні бібліотеки інженерних функцій до даних, щоб здійснити бажані перетворення. Результатом має бути набір даних, організований для досягнення оптимального балансу між часом навчання для нової моделі та необхідним обчисленням;

д) перевірка даних. На цьому етапі дані розбиваються на два набори. Перший набір використовують для навчання моделі машинного навчання або глибокого навчання. Другий набір – це дані тестування, які використовуються для оцінки точності та надійності отриманої моделі. Цей другий крок допомагає виявити будь-які проблеми в гіпотезі, що використовується при очищенні та проектуванні ознак даних. Якщо вчені, які працюють із даними, задоволені результатами, вони можуть передати завдання попередньої обробки інженеру з обробки даних, який вигадає, як масштабувати його для виробництва. Якщо ні, фахівці за даними можуть повернутися та внести зміни до способу реалізації етапів очищення даних та проектування функцій.

Як зазначалося раніше, попередня обробка даних відіграє ключову роль на ранніх етапах машинного навчання та розробки додатків ШІ. У контексті штучного інтелекту попередня обробка даних використовується для покращення способу очищення, трансформації та структурування даних, щоб підвищити точність нової моделі, одночасно зменшуючи обсяг необхідних обчислень.

Хороший конвеєр попередньої обробки даних може створювати повторно використовувані компоненти, які спрощують тестування різних ідей для оптимізації бізнес-процесів або підвищення рівня задоволеності клієнтів. Наприклад, попередня обробка може покращити спосіб організації даних для механізму рекомендацій шляхом вдосконалення вікових діапазонів, які використовуються для категоризації клієнтів [35].

2.4 Метрики оцінки якості регресійної моделі

Оцінка якості регресійних моделей дозволяє нам кількісно визначити, наскільки добре наші моделі узагальнюють невідомі дані, виявити потенційні проблеми, як-от переобладнання або недостатнє пристосування, і вибрати найкращу модель із кількох кандидатів. Це важливий крок у робочому процесі обробки даних, який часто робить різницю між моделлю, яка надає цінну інформацію, і моделлю, яка вводить нас в оману.

Оцінка відіграє вирішальну роль у створенні ефективних регресійних моделей. Визначаючи кількісно продуктивність моделі, метрики оцінки забезпечують відчутне вимірювання того, наскільки добре працює модель. Ці показники дають можливість порівняти різні моделі та вибрати ту, яка працює найкраще. Без показників оцінки ми залишаємося на припущеннях і інтуїції. Метрики оцінювання усувають цю неоднозначність. Вони дозволяють ранжувати моделі на основі їх продуктивності та вибирати ту, яка дає найточніші прогнози.

Окрім вибору моделі, оцінка також допомагає налаштувати обрані моделі. Налаштовуючи параметри та відстежуючи, як ці зміни впливають на показники оцінки, можливо оптимізувати наші моделі та підвищити їх ефективність.

Нездатність правильно оцінити регресійні моделі може призвести до кількох пасток. Погано оцінена модель може чудово працювати з вашими навчальними даними, але з тріском зазнає невдачі, коли зіткнеться з невидимими даними. Цей сценарій, відомий як переобладнання, є поширеним наслідком нехтування належною оцінкою моделі. Зосереджуючись виключно на навчальних даних і не враховуючи, як модель узагальнюється, ми ризикуємо розробити моделі, які є надто складними та не в змозі вловити основний шаблон.

З іншого боку, ми маємо недостатню підгонку, коли наша модель надто проста, щоб охопити базовий шаблон навіть у навчальних даних. Без належної оцінки ми можемо пропустити цю проблему, що призведе до моделі з низькою

продуктивністю. Без належної оцінки також легко потрапити в пастку «парадоксу точності» – модель, упереджена до прогнозування більшості класів, може мати високий рівень точності, але це не означає, що модель хороша. Наприклад, якщо 95% електронних листів не є спамом, модель, яка завжди передбачає «не спам», буде точною на 95%, але вона не зможе вловити жодного фактичного спама.

Належне оцінювання захищає від цих пасток, спрямовуючи нас до моделей, які не тільки добре працюють на наших навчальних даних, але й добре узагальнюють невидимі дані. Коли ми орієнтуємося в складному ландшафті регресійного моделювання, показники оцінювання служать нашою північною зіркою, спрямовуючи нас до ефективних і надійних моделей.

Перш ніж ми заглибимося в специфіку кожного показника, важливо зрозуміти, чого ці показники в основному намагаються досягти. По суті, вони кількісно визначають різницю між прогнозованими значеннями, які генерує наша модель, і фактичними значеннями, які ми спостерігали в даних. Чим менше ця різниця, тим краще працює наша модель.

Проте не всі відмінності розглядаються однаково. Деякі показники більше зосереджуються на більших відмінностях, караючи моделі більше, якщо вони роблять кілька великих помилок. Інші ставляться до всіх відмінностей однаково, незалежно від їх розміру. Розуміння цих нюансів має вирішальне значення під час вибору та інтерпретації показників оцінювання.

Далі розглянемо п'ять ключових показників, які використовуються для оцінки регресійних моделей:

- R-квадрат (R^2): це, мабуть, найвідоміший показник для регресійних моделей. Він вимірює частку загальної варіації залежної змінної, яку фіксує модель;

- скоригований R-квадрат: це модифікована версія R-квадрату, яка була скоригована відповідно до кількості предикторів у моделі. Воно збільшується, лише якщо новий предиктор покращує модель більше, ніж можна було очікувати випадково;

– середня квадратична помилка (MSE): це середнє значення квадратів різниць між прогнозованими та фактичними значеннями. Це надає більшої ваги більшим відмінностям і особливо корисно, коли ми маємо несподівані значення, які ми хочемо взяти до уваги;

– середньоквадратична помилка (RMSE): це квадратний корінь із MSE. Витягуючи квадратний корінь MSE, ми повертаємо метрику помилки до тієї ж одиниці, що й цільова змінна, що часто полегшує інтерпретацію;

– середня абсолютна похибка (MAE): це середнє значення абсолютних різниць між прогнозованими та фактичними значеннями. На відміну від MSE, він розглядає всі відмінності однаково і менш чутливий до викидів.

Надалі глибше зануримося в кожен з цих моделей.

Показник 1. R-квадрат (R^2).

R-квадрат, також відомий як коефіцієнт детермінації, є одним з показників, що найчастіше використовуються для оцінки відповідності регресійної моделі. Він забезпечує міру того, наскільки результати, що добре спостерігаються, відтворюються моделлю, на основі частки загальної варіації результатів, поясненої моделлю. R-квадрат, що дорівнює 100%, вказує на те, що всі зміни залежної змінної повністю пояснюються змінами незалежної змінної.

R-квадрат визначається як частка дисперсії залежної змінної, яку можна передбачити незалежною змінною(ям). Математично це можна розрахувати як

$$R^2 = 1 - (SSR / SST),$$

де SSR – це сума квадратів залишків. Залишки є різницею між фактичними значеннями залежної змінної і прогнозованими значеннями регресійної моделі ;

SST (загальна сума квадратів) – це загальна сума квадратів, яка є сумою різниць між фактичними значеннями залежної змінної і середнім значенням залежної змінної, зведеними в квадрат.

Значення R^2 знаходиться в діапазоні від 0 до 1. Значення 1 означає, що модель точно передбачає залежну змінну, використовуючи незалежну

змінну (i). Значення 0 означає, що модель взагалі може передбачити залежну змінну, використовуючи незалежну змінну.

Інтерпретація R-квадрату проста. Це десяткове значення від 0 до 1, яке під час обговорення відповідності моделі часто виявляється у відсотках. R-квадрат, що дорівнює 100% (або 1, якщо не помножено на 100), вказує на те, що всі зміни залежної змінної повністю пояснюються змінами незалежної змінної. Іншими словами, наша модель ідеально відповідає даним. З іншого боку, R-квадрат, що дорівнює 0%, вказує на те, що залежну змінну взагалі неможливо передбачити на основі незалежної змінної.

Незважаючи на широке використання та простоту інтерпретації, R^2 має свої обмеження:

Він вимірює лише пояснену дисперсію: R-квадрат не повідомляє нам, чи є обрана модель хорошою чи поганою, і не передає надійність моделі. Він лише кількісно визначає величину мінливості цільової змінної, яка враховується предикторами в моделі:

– чутливість до непотрібних функцій: значення R-квадрат або залишиться незмінним, або збільшиться з додаванням інших змінних, навіть якщо ці змінні лише слабо пов'язані з відповіддю. Це може призвести до переобладнання, особливо коли йдеться про багато функцій;

– не підходить для порівняння різних наборів даних: R-квадрат не є хорошим показником для порівняння продуктивності моделі в різних наборах даних. Оскільки він вимірює частку дисперсії, її значення може суттєво відрізнятись залежно від варіації дисперсії різних наборів даних;

– нижча продуктивність із нелінійними даними: хоча R-квадрат є хорошим показником для лінійної регресії, він не так добре працює з нелінійними шаблонами даних.

Розуміння цих обмежень має вирішальне значення для правильного використання та інтерпретації R^2 . У наступних розділах ми розглянемо додаткові метрики оцінки, які можуть доповнити R^2 та надати повнішу картину ефективності моделі.

Показник 2. Скоригований R-квадрат.

Скоригований R-квадрат – це модифікована версія R-квадрату, яка була скоригована з урахуванням кількості предикторів моделі. Як і R-квадрат, він забезпечує міру частки загальної дисперсії залежною змінною, яка пояснюється незалежними змінними. Однак він також враховує кількість використовуваних предикторів, додаючи штраф за складність моделі.

Хоча R-квадрат і скоригований R-квадрат дозволяють оцінити, наскільки добре модель відповідає даним, між ними є принципова різниця: R-квадрат передбачає, що кожна окрема змінна пояснює зміну залежної змінної, тоді як скоригований R-квадрат додає штраф за непотрібну складність моделі. Проблема з R^2 полягає в тому, що він має тенденцію переоцінювати продуктивність моделі з додаванням додаткових змінних, навіть якщо ці змінні лише слабо пов'язані з відповіддю. Це може призвести до переоснащення та помилково високих значень R-квадрату. Скоригований R-квадрат вирішує цю проблему, зменшуючи значення, коли модель включені непотрібні предиктори. Це робить скоригований R-квадрат надійнішим заходом для оцінки загальної якості регресійної моделі, особливо при порівнянні моделей з різною кількістю предикторів.

Коригований R-квадрат слід використовувати, якщо у вас є кілька моделей регресії з різною кількістю предикторів. У міру того, як ви додаєте до моделі більше предикторів, R-квадрат завжди буде збільшуватися, навіть якщо ці предиктори не приносять реального покращення моделі. Це може ускладнити визначення того, чи пов'язане збільшення R-квадрату з дійсно кращою моделлю або просто зі збільшенням її складності. Навпаки, скоригований R-квадрат збільшується тільки в тому випадку, якщо нова змінна покращує модель більше, ніж можна було б очікувати випадково, і зменшується, коли предиктор покращує модель менше, ніж очікувалося випадково. Це робить його надійнішою метрикою при порівнянні моделей різної складності.

Як і R-квадрат, скоригований R-квадрат є десятковим числом від 0 до 1 і часто виражається у відсотках. Скоригований R-квадрат, що дорівнює 100%,

вказує на те, що всі зміни залежної змінної повністю враховуються незалежними змінними моделі. Оцінка 0% свідчить про те, що модель не пояснює жодних відхилень даних відповіді середнього значення. Однак, на відміну від R-квадрату, скоригований R-квадрат враховує кількість предикторів моделі. Наприклад, якщо у вас є дві моделі з однаковим R-квадратом, але різною кількістю предикторів, модель з меншою кількістю предикторів матиме вищий скоригований R-квадрат, що відображає той факт, що вона досягла однакового ступеня відповідності з меншою кількістю предикторів. Інтерпретація скоригованого R-квадрату ізольовано може ввести в оману, оскільки не дає інформації про те, яку частину незрозумілої дисперсії можна було б пояснити за допомогою кращої моделі. Як і у випадку з будь-якою іншою метрикою, завжди краще використовувати скоригований R-квадрат разом з іншими метриками, щоб отримати більш повну картину продуктивності вашої моделі.

Показник 3. Середня квадратична помилка (MSE).

Середньоквадратична помилка (MSE) – ще один показник, який використовується для оцінки ефективності регресійних моделей. На відміну від R-квадрату та скоригованого R-квадрату, які є мірами поясненої дисперсії, MSE є мірою помилки прогнозування. Зокрема, він визначає середньоквадратичну різницю між фактичними та прогнозованими значеннями. MSE дає вищий штраф за великі помилки, зводячи залишки в квадрат. Це означає, що моделі, які роблять великі помилки, призведуть до більшого MSE. Таким чином, при використанні MSE як показник наша мета – мінімізувати його значення.

MSE оцінювача вимірює середнє значення квадратів помилок, тобто середню квадратичну різницю між оціненими значеннями та фактичним значенням. Математично це визначається як

$$MSE = \frac{1}{n} * \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 ,$$

де n – кількість спостережень;

Y_i – фактичне значення залежної змінної для i -го спостереження;

\hat{Y}_i – прогнозоване значення залежної змінної для i -го спостереження.

MSE завжди є невід'ємним, а значення 0 вказує на ідеальну відповідність даним. На практиці це рідко трапляється поза сценаріями переобладнання.

MSE дає нам абсолютну міру величини помилки моделі. Він каже нам, наскільки точно передбачення моделі відповідають спостережуваним даним. Чим менше MSE, тим ближча відповідність даним. Однак одна з проблем інтерпретації MSE полягає в тому, що вона не одразу стає інтуїтивно зрозумілою. Він не говорить нам прямо, наскільки наш прогноз у середньому відхилятиметься від фактичного значення. Натомість він повідомляє нам, що в середньому наш прогноз буде відхилятися від фактичного значення на квадратний корінь із MSE. Інша проблема полягає в тому, що MSE залежить від масштабу. Це означає, що важко порівнювати MSE моделей для різних наборів даних, якщо ці набори даних не вимірюють одну і ту ж змінну в одному масштабі.

MSE має свої плюси і мінуси:

Плюси MSE:

– підкреслює більші помилки: шляхом зведення залишків у квадрат MSE надає більшу вагу більшим помилкам. Це може бути корисним, коли більші помилки особливо небажані;

– диференційність: квадратична функція має похідні, що робить MSE більш зручним для оптимізації в алгоритмах машинного навчання.

Мінуси MSE:

– чутливий до викидів: оскільки MSE зводить залишки у квадрат, він може бути дуже чутливим до викидів. Один викид потенційно може мати великий вплив на MSE;

– залежно від масштабу: MSE залежить від масштабу, що означає, що ви не можете порівнювати MSE різних змінних, які знаходяться в різних масштабах;

– без прямої інтерпретації: одиниці MSE не збігаються з одиницями цільової змінної. Це ускладнює інтерпретацію в діловому середовищі.

Показник 4. Середньоквадратична помилка (RMSE).

RMSE – гарний захід, який можна використовувати, коли більше дбаєте про покарання за великі помилки. Зводячи помилки в квадрат перед їх усередненням, RMSE надає більшої ваги більшим помилкам. Це означає, що RMSE є найбільш корисним у контекстах, де великі помилки особливо небажані. Як і MSE, RMSE зазвичай використовується в регресійному аналізі та прогнозуванні, де метою часто є мінімізація великих помилок. Це також зручний показник, який можна використовувати, коли ви хочете пояснити продуктивність моделі більш зрозумілим способом, оскільки її одиниці виміру збігаються з цільовою змінною.

RMSE вимірює середню величину помилки. Він повідомляє вам, яку помилку система зазвичай припускається у своїх прогнозах, з вищою вагою для великих помилок. Значення RMSE інтерпретується в тих самих одиницях, що й змінна відповіді, що спрощує зв'язок із прогнозованою змінною. Найменше значення RMSE буде вказувати на кращу відповідність даним, тоді як більше значення вказує на найгіршу відповідність. Однак RMSE, що дорівнює нулю, не завжди є кінцевою метою. Модель, яка надто ідеально відповідає навчальним даним, може призвести до перенавчання, коли модель добре працює на навчальних даних, але погано на нових, невидимих даних. Як і у випадку з будь-якою метрикою, важливо використовувати RMSE у поєднанні з іншими метриками, щоб отримати повне уявлення про продуктивність вашої моделі.

Показник 5. Середня абсолютна похибка (MAE).

Середня абсолютна помилка (MAE) – це ще один показник, який використовується для вимірювання продуктивності регресійної моделі. Подібно до MSE і RMSE, він кількісно визначає різницю між фактичними і прогнозованими значеннями. Однак, на відміну від MSE та RMSE, які зводять залишки у квадрат, MAE набуває абсолютного значення цих різниць. Приймаючи абсолютне значення, MAE надає однакову вагу всім помилкам,

незалежно від того, великі вони чи малі. Це робить MAE відмінним показником, який можна використовувати, коли ви хочете дізнатися про середню величину помилки, але не хочете надмірно карати за великі помилки.

Математично MAE визначається як середнє значення абсолютних різниць між прогнозованими та фактичними значеннями. Він розраховується таким чином:

$$MAE = \frac{1}{n} * \sum_{i=1}^n |Y_i - \hat{Y}_i|,$$

де n – кількість спостережень;

Y_i – фактичне значення залежної змінної для i -го спостереження;

\hat{Y}_i – прогнозоване значення залежної змінної для i -го спостереження.

MAE завжди є невід'ємним, при цьому нижче значення вказує на кращу відповідність даним.

MAE дає уявлення про те, наскільки хибними були прогнози в середньому. Як і RMSE, воно вимірюється в тих самих одиницях, що і цільова змінна, тому його відносно легко інтерпретувати. Найменше значення MAE вказує на кращу відповідність моделі даних. MAE, що дорівнює 0, означає, що модель дає точні прогнози (що практично малоімовірно, якщо ви не переоснащуєте свою модель). Один із способів інтерпретації MAE полягає в наступному: враховуючи MAE, що дорівнює « x », ви можете сказати, що в середньому ваш прогноз не відповідає істинному значенню на « x » одиниць.

Як і будь-який інший показник, MAE має свої плюси та мінуси.

Плюси MAE:

- легко зрозуміти та розрахувати: MAE легко зрозуміти та розрахувати. Він забезпечує простий спосіб уявлення середньої помилки;

- менш чутливий до викидів: оскільки MAE не зводить у квадрат залишки, він менш чутливий до викидів порівняно з MSE та RMSE. Це робить його найкращим показником, коли викиди не викликають особливого занепокоєння.

Мінуси MAE:

– відсутність акценту великих помилках. Хоча менша чутливість до викидів може бути перевагою, вона може бути недоліком, коли великі помилки особливо небажані;

– чи не диференційовано в нулі. На відміну від MSE та RMSE, MAE не диференціюється в нулі, що робить його менш придатним для певних алгоритмів машинного навчання, заснованих на диференціюванні.

Порівняємо метрики:

а) R-квадрат і скоригований R-квадрат: обидва ці показники дозволяють оцінити, яка частина мінливості залежної змінної пояснюється незалежними змінними. Однак вони не дають нам жодної інформації про абсолютний розмір помилки. Скоригований R-квадрат має додаткову перевагу, яка полягає в обліку кількості предикторів моделі, що допомагає уникнути перенавчання;

б) MSE та RMSE: I MSE, і RMSE надають більшої ваги більшим помилкам шляхом зведення залишків у квадрат. Вони корисні, коли великі помилки є особливо небажаними. Ключова різниця між ними полягає в тому, що RMSE вимірюється в тих самих одиницях, що і залежна змінна, що спрощує інтерпретацію. Як MSE, і RMSE можуть сильно залежати від викидів;

в) MAE: На відміну від MSE та RMSE, MAE однаково обробляє всі помилки, приймаючи абсолютне значення залишків. Він забезпечує чітке уявлення середньої помилки і менш чутливий до викидів. Однак він не приділяє особливої уваги великим помилкам.

Загалом, жоден показник не є «найкращим» у всіх ситуаціях. Вибір того, який показник використовувати, залежить від конкретного контексту, наявності викидів, важливості більших помилок і потреб зацікавлених сторін.

Кожна з цих метрик пропонує унікальний погляд на продуктивність вашої моделі. R-квадрат і скоригований R-квадрат дають уявлення про те, яку частину цільової дисперсії змінної може пояснити ваша модель. MSE та RMSE дозволяють виміряти середню помилку прогнозу, при цьому RMSE надає

більшим помилкам більшої ваги. А MAE дає уявлення про середню абсолютну помилку, не надаючи додаткової ваги великим помилкам [38].

Висновки за розділом 2

У цьому дослідженні вибір комбінації згорткових та рекурентних нейронних мереж (CNN-RNN) для розв'язання задачі розпізнавання емоцій на відео є стратегічно обґрунтованим і визначається кількома ключовими факторами. Згорткові нейронні мережі (CNN) добре справляються з аналізом візуальних даних, витягаючи просторові ознаки з кадрів, тоді як рекурентні нейронні мережі (RNN) додають здатність аналізувати часові залежності в динаміці емоційних виразів.

Важливо відзначити, що комбінація цих двох типів мереж створює універсальний підхід до розпізнавання емоцій, дозволяючи моделі адаптуватися до різноманітних виразів та контекстів відеоспостереження. Здатність CNN-RNN адекватно враховувати еволюцію емоцій у часі робить його особливо корисним для аналізу динамічних сценаріїв, де емоції можуть змінюватися від кадру до кадру.

Крім того, вибір такої архітектури має практичний аспект, оскільки цей підхід показує високу ефективність на великих наборах відеоданих, таких як Aff-Wild2. Комплексне використання згорткових та рекурентних нейронних мереж дозволяє досягнути оптимального балансу між просторовим та часовим аналізом для задачі розпізнавання емоцій на відео.

Отже, вибір CNN-RNN архітектури вказує на високий потенціал цього підходу в контексті розпізнавання емоцій на відео та його адаптабельність до різноманітних сценаріїв та умов спостереження.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Високорівнева мова програмування Python

Python – це високорівнева мова програмування з відкритим вихідним кодом та крос-платформною підтримкою, яка набула популярності протягом останніх десяти років. Вона була випущена вперше у 1991 році, а її остання версія – 3.7.0. CPython є референтною реалізацією мови програмування Python та вважається її стандартною та широко використовуваною реалізацією. Реалізована на мові програмування C, CPython – це основна та найбільш поширена реалізація мови.

Мова програмування Python була створена Гвідо ван Россумом, голландським програмістом. Ван Россум працював у центрі математики та інформатики в Нідерландах, і він шукав мову програмування, яка була б простою у використанні, але також потужною та гнучкою.

Одним із ключових моментів є легкість вивчення та розуміння синтаксису Python. Це особливо важливо, оскільки реалізація системи розпізнавання емоцій вимагає глибокого розуміння інструментів та алгоритмів, що використовуються. Python робить цей процес більш доступним та приємним.

Сьогодні Python є однією з найпопулярніших мов програмування у світі. Вона використовується в широкому спектрі застосунків, включаючи веб-розробку, машинне навчання, штучний інтелект та наукові розрахунки.

3.2 Алгоритм розв’язання задачі розпізнавання емоцій

Побудуємо алгоритм розв’язання задачі оцінки емоційних характеристик людини таких, як емоційна толерантність та збудженість з відеопотоку з обличчям людини, його можна розбити на три етапи, такі як:

– збір даних;

– побудова архітектури моделі з програмною реалізацією її навчання та оцінки якості;

– оцінка якості моделі на основі проведених експериментів.

На рисунку 3.1 можна побачити алгоритм розв'язання задачі розпізнавання емоцій.

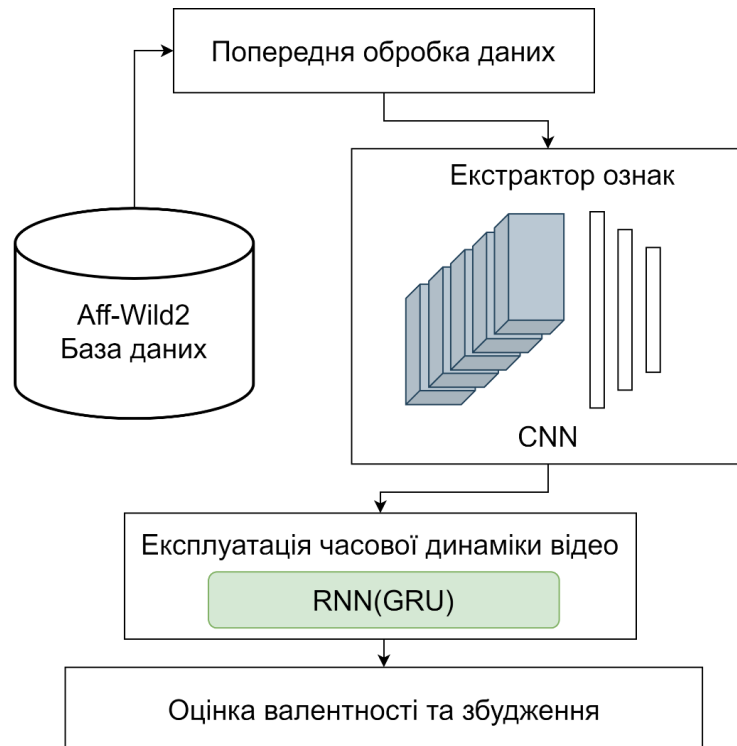


Рисунок 3.2 – Програмна реалізація рекурентної нейронної мережі (RNN)

На першому етапі збирається набір даних для навчання та тестування нейронної мережі. В нашому випадку використовується база даних Aff-Wild2 [39].

Набір даних Aff-Wild2 є розширенням набору даних Aff-Wild для розпізнавання афектів. Він приблизно вдвічі збільшує кількість включених відеокадрів і кількість суб'єктів, таким чином покращуючи варіативність включених поведінкових реакцій і залучених осіб.

Набір даних Aff-Wild2 складається з 564 відеороликів, що містять близько 2,8 млн кадрів, з 554 суб'єктами (326 з яких – чоловіки і 228 – жінки).

Всі відео були анотовані з точки зору валентності та збудження. 546 відео з приблизно 2,6 млн кадрів було анотовано з точки зору основних висловлювань. 541 відео з приблизно 2,6 млн кадрів було анотовано з точки зору одиниць дії. Набір даних Aff-Wild2 демонструє велику різноманітність за віком, етнічною та національною приналежністю суб'єктів. Він також має великі варіації та різноманітність середовищ.

На другому етапі відбувається побудова архітектури моделі з програмною реалізацією її навчання та оцінки якості. Для цього потрібно побудувати архітектуру згорткової нейронної мережі, яка буде використовуватися для вилучення ознак з зображень відеоряду, також необхідно побудувати архітектуру рекурентної мережі, яка буде використовуватися для отримання кінцевих прогнозувань щодо вказаних вище емоційних характеристик. Необхідно імплементувати у вигляді програмного коду архітектури нейронних мереж та забезпечити реалізацію їх навчання та оцінки якості роботи.

На третьому етапі відбувається оцінка якості моделі на основі проведених експериментів. Потрібно визначити які метрики оцінки будуть найкраще відповідати вимогам задачі, проаналізувати їх з точки зору математичної основи, та розробити програмну реалізацію для підрахунку їх значень протягом навчання та тестування.

У результаті роботи даного алгоритму буде побудована модель, яка з високою точністю зможе розпізнавати емоції людини. На рисунках 3.2, 3.3 зображено програмну реалізацію екстрактору ознак (CNN) та рекурентної нейронної мережі (RNN).

```

def simple_cnn_ccc_loss(self):
    model = Sequential()
    model.add(BatchNormalization(input_shape=self.input_shape))
    model.add(Conv2D(64,
                    (5, 5),
                    padding='same',
                    activation='relu',
                    use_bias=False))
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Conv2D(128,
                    (5, 5),
                    activation='relu',
                    use_bias=False))
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Conv2D(256,
                    (5, 5),
                    activation='relu',
                    use_bias=False))
    model.add(AvgPool2D(pool_size=(9, 9),
                        strides=9))
    model.add(Flatten())
    model.add(Dense(300))
    model.add(Dropout(0.5))
    model.add(Dense(2))

    lr_schedule = ExponentialDecay(
        initial_learning_rate=0.01,
        decay_steps=10000,
        decay_rate=0.9)
    opt = SGD(learning_rate=lr_schedule, momentum=.9)
    model.compile(loss=metrics.ccc_loss,
                  optimizer=opt,
                  metrics=[metrics.rmse,
                          metrics.rmse_v,
                          metrics.rmse_a,
                          metrics.cc_v,
                          metrics.cc_a,
                          metrics.ccc_v,
                          metrics.ccc_a])

    return model

```

Рисунок 3.2 – Програмна реалізація екстрактору ознак (CNN)

```

def gru_ccc_loss(self):
    model = Sequential()
    model.add(GRU(100, return_sequences=True, input_shape=[100, 300]))
    model.add(GRU(100, return_sequences=True))
    model.add(GRU(50, return_sequences=False))
    model.add(Dense(2))

    lr_schedule = ExponentialDecay(
        initial_learning_rate=1e-2,
        decay_steps=400,
        decay_rate=0.97)
    opt = Adam(learning_rate=lr_schedule)
    # opt = SGD(lr=.01, decay=1e-5, momentum=.9)
    model.compile(loss=metrics.ccc_loss,
                  optimizer=opt,
                  metrics=[metrics.rmse,
                          metrics.rmse_v,
                          metrics.rmse_a,
                          metrics.cc_v,
                          metrics.cc_a,
                          metrics.ccc_v,
                          metrics.ccc_a])

    return model

```

Рисунок 3.3 – Програмна реалізація рекурентної нейронної мережі (RNN)

3.3 Опис програми

Програма для вирішення задач розпізнавання людських емоцій реалізована на мові програмування Python. Дана програма робить оцінку емоційних характеристик (емоційна толерантність та збудженість) на основі відеозображення облич людей.

Метою даної програми є побудова архітектур двох нейронних мереж які утворюють комбінацію CNN-RNN та їх навчання з подальшою оцінкою якості передбачень.

Для тренування нейронних мереж використали.

Програма містить реалізацію Tensorflow Keras підходу для розпізнавання емоцій на великому наборі відеоданих на основі згорткового екстрактора ознак і рекурентної нейронної мережі, який навчається та тестується наборі даних Aff-Wild2.

Ми надаємо моделі, навчені на Aff-Wild2, в папку з чекпоїнтами і логи тензорної дошки під час навчання цих моделей в папку з логами. Також всі експерименти, такі як валідація CNN на невеликих наборах даних, візуалізація шарів CNN, навчання та прогнозування наведені в коді.

Для використання програми по-перше потрібно підготувати віртуальне середовище, клонуванням репозиторію з програмним кодом та встановленням необхідних залежностей в віртуальне середовище.

По-друге, необхідно виконати процес тренування:

а) вказати конфігурацію для навчання в файлі `config.py`.

Цей код містить конфігураційні налаштування для різних середовищ, таких як "colab", "local" та "vm". Залежно від обраного середовища, змінні, які вказують на шляхи до даних, каталоги з контрольними точками моделі та інші параметри, будуть налаштовані відповідно;

б) за допомогою терміналу з області віртуального середовища запустити тренувальний процес згорткової нейронної мережі за допомогою файлу `train_cnn.py`;

Даний скрипт призначений для тренування моделей згорткових нейронних мереж (CNN) на кадрах. Він використовує бібліотеки TensorFlow та Keras для реалізації процесу навчання. Скрипт визначає функції для створення завантажувачів даних – `train_generator` та `valid_generator`, і тренування моделі – `train_model`. Основна функція `main` встановлює основні параметри, створює зворотні виклики для збереження кращих моделей та збереження в TensorBoard, ініціалізує модель дослідження за вказаною архітектурою. Наприкінці отримує генератори даних і починає процес тренування на 100 епох з вказаними зворотними викликами. Скрипт також перевіряє наявність каталогів і забезпечує створення необхідних папок для збереження кращих моделей та журналів;

в) використати навчену згорткову нейронну мережу для вилучення ознак з кадрів відеоряду за допомогою файлу `extract_features.py`;

Даний скрипт генерує вилучені ознаки для кожного відео, які можуть використовуватися іншими моделями. Він використовує попередньо навчену модель вилучення ознак на основі згорткової нейронної мережі для обробки кадрів та вилучення відповідних ознак. Скрипт проходить через тренувальний та тестовий набори даних, вилучаючи ознаки для кожного відео та зберігаючи їх у відповідному каталозі оброблених послідовностей. Він ефективно перевіряє, чи вже існують вилучені ознаки для конкретного відео, щоб уникнути зайвого обчислення. Процес вилучення ознак відображається з використанням `tqdm` для візуального показу прогресу. Загалом скрипт розроблено для створення послідовностей ознак, які можуть бути використані іншими моделями в обробці відео;

г) запустити процес тренування рекурентної нейронної мережі на попередньо створених мапах ознак за допомогою файлу `train_rnn.py`.

Наданий код розроблено для тренування різних рекурентних нейронних мереж (RNN) на збережених векторах ознак. Сценарій використовує бібліотеку TensorFlow та Keras для навчання моделей, використовуючи збережені вектори ознак замість обробки вхідних кадрів. У коді реалізовані функції для отримання

генераторів даних – `get_generators`, тренування моделі – `train_model`, та основна функція `main`, яка налаштовує параметри, створює зворотні виклики для збереження кращих моделей та ведення журналів в TensorBoard. Можна вибрати різні архітектури RNN, такі як GRU або простий RNN, та вказати кількість епох тренування. Результати навчання зберігаються у відповідному каталозі для подальшого використання.

По-третє, потрібно виконати оцінку роботи моделі, виконанням коду з файлу `predict.py`.

Цей код використовує збережені моделі згорткових нейронних мереж (CNN) та рекурентних нейронних мереж (RNN), щоб здійснювати прогнози для відео. Здійснюється підготовка та обробка послідовностей ознак з кадрів відео, використовуючи витягнуті ознаки з CNN та подані на вхід у RNN для отримання прогнозів емоційних оцінок (валентності та збудженості). Результати прогнозів зберігаються у відповідний каталог. Код також враховує випадок, коли кількість кадрів менше `RNN_WINDOW_SIZE`, та можливість використання попередніх прогнозів за допомогою CNN для початкових фреймів. Усі етапи виконуються для відео з приватного тестового набору даних.

Висновки за розділом 3

Представлений опис програми розпізнавання людських емоцій на відео базується на комбінації згорткових та рекурентних нейронних мереж (CNN-RNN). Основна мета - навчання та тестування моделей на датасеті Aff-Wild2 для отримання точних прогнозів емоційних характеристик.

Програма використовує попередньо навчені моделі для вилучення ознак з відеокадрів та робить прогнози для емоційної толерантності та збудженості. Описані етапи включають тренування, генерацію вилучених ознак та оцінку якості за допомогою метрик.

Потрібно підготувати віртуальне середовище, налаштувати конфігурацію для навчання, виконати тренування та використовувати навчені моделі для прогнозування емоцій. Програма дозволяє ефективно поєднати згорткові та рекурентні мережі для кращої точності розпізнавання емоцій на відео.

4 РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ ТА ЇХ АНАЛІЗ

4.1 Навчання та оцінка якості роботи моделей CNN та RNN

В рамках експерименту було проведено навчання моделі згорткової нейронної мережі. Були дістані ознаки з зображень набору даних за допомогою навченої згорткової нейронної мережі та потім на цих даних була натренована рекурентна нейронна мережа. В кінці експерименту були виведені метрики оцінки якості роботи моделі, такі як:

- `ccc_loss` – функція втрат, яка використовує дві інші метрики – `ccc_v` та `ccc_a` (Concordance Correlation Coefficient для емоційної толерантності та збудженості відповідно). Вона обчислює середнє значення між `ccc_v` та `ccc_a` і віднімає його від 1;

- `gmse_v` – обчислює квадратний корінь середнього квадрату різниці між прогнозованими та справжніми значеннями для емоційної толерантності;

- `gmse_a` – обчислює квадратний корінь середнього квадрату різниці між прогнозованими та справжніми значеннями для збудженості;

- `gmse` – обчислює квадратний корінь середнього квадрату різниці між прогнозованими та справжніми значеннями;

- `cc_v` – обчислює коефіцієнт кореляції Пірсона між прогнозованими та справжніми значеннями для емоційної толерантності;

- `cc_a` – обчислює коефіцієнт кореляції Пірсона між прогнозованими та справжніми значеннями для збудженості;

- `ccc_v` – обчислює Concordance Correlation Coefficient для емоційної толерантності, який враховує кореляцію, стандартне відхилення та середню відстань між прогнозованими та справжніми значеннями;

- `ccc_a` – обчислює Concordance Correlation Coefficient для збудженості, який враховує кореляцію, стандартне відхилення та середню відстань між прогнозованими та справжніми значеннями.

Експеримент проводився з наступними налаштуваннями для параметрів:

а) згорткова нейронна мережа:

- 1) steps_per_epoch = 1000 – кількість ітерацій за одну епоху;
- 2) nb_epoch = 100 – кількість епох;
- 3) dim = (96, 96) – розмірність вхідного зображення;
- 4) batch_size = 128 – розмір батчу;

б) рекурентна нейронна мережа:

- 1) nb_epoch = 100 – кількість епох;
- 2) batch_size = 256 – розмір батчу.

Згорткова нейронна мережа була натренована для вилучення ознак, та була використана для їх вилучення з набору даних. Вилучені ознаки були використані для навчання рекурентної нейронної мережі.

В таблицях 4.1, 4.2 наведені значення метрик для тренувального набору та валідаційного для рекурентної нейронної мережі.

Таблиця 4.1 – Значення метрик для тренувального набору даних для рекурентної нейронної мережі

сс_a	сс_v	ccc_a	ccc_v	loss	rmse	rmse_a	rmse_v
0.9474	0.9605	0.9466	0.96	0.0467	0.1051	0.0967	0.1135

Таблиця 4.2 – Значення метрик для валідаційного набору даних для рекурентної нейронної мережі

сс_a	сс_v	ccc_a	ccc_v	loss	rmse	rmse_a	rmse_v
0.8251	0.85	0.8201	0.85	0.3	0.743	0.832	0.913

Висновки за розділом 4

Експеримент включав тренування згорткової нейронної мережі для вилучення ознак і їх використання в рекурентній нейронній мережі для емоційної толерантності та збудженості.

Отримані результати для тренувального та валідаційного наборів показують високі значення коефіцієнтів кореляції Пірсона (cc_a та cc_v), коефіцієнтів кореляції конкордації (ccc_a та ccc_v) та низькі значення середньоквадратичної похибки, що свідчить про ефективність моделі в прогнозуванні емоційної толерантності та збудженості. Модель демонструє низькі втрати (loss), що підтверджує її високу точність та адаптивність до тренувальних та валідаційних даних.

Таким чином, отримані результати свідчать про успішність використання комбінації згорткової та рекурентної нейронних мереж для аналізу емоційної толерантності та збудженості на зображеннях, з високим потенціалом застосування в області розпізнавання емоційних станів.

ВИСНОВКИ

У ході даної кваліфікаційної роботи було побудовано та удосконалено модель розпізнавання емоцій на основі відеофрагмента обличчя. Актуальність даного дослідження визначається зростанням кількості магазинів та публічних просторів, де спостереження за емоційним станом людей стає важливою задачею. Використання штучного інтелекту в даному контексті дозволяє аналізувати величезні обсяги відео-даних у реальному часі, перевищуючи можливості людського спостереження.

Було проведено тренування згорткової нейронної мережі для вилучення ознак та їх використання в рекурентній нейронній мережі для емоційної толерантності та збудженості. Результати дослідження на тренувальному та валідаційному наборах показали високі значення коефіцієнтів кореляції та конкордації, а також низькі втрати, свідчачи про ефективність моделі в прогнозуванні емоційної толерантності та збудженості.

Отримані результати свідчать про успішність використання комбінації згорткової та рекурентної нейронних мереж для аналізу емоційної толерантності та збудженості на зображеннях. Модель має великий потенціал застосування в різних галузях, включаючи безпеку, медицину та маркетинг, демонструючи високу точність та адаптивність до реальних умов. Такий підхід до розпізнавання емоцій може сприяти покращенню технологічних, психологічних та медичних аспектів взаємодії людини з оточуючим середовищем.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Yuvchenko K., Yesilevskyi V., Sereda O. Human emotion recognition system using deep learning algorithms. *Innovate technologies and scientific for industries*. 2022. № 3. С. 60-69.
2. Ювченко К. С, Єсілевський В. С., Серeda О. Г. Розпізнавання емоцій за допомогою нейронного класифікатора. *Міжнародна науково-практична конференція «Інформаційні технології та комп'ютерне моделювання»* : матеріали конференції (м. Івано-Франківськ, 15–16 грудня 2022 року). С. 41-43.
3. Ювченко К. С, Єсілевський В. С., Серeda О. Г. Класифікатор емоцій людини за зображенням обличчя на основі нейронної мережі. *11-а Міжнародна науково-технічна конференція «Information Systems and Technology» (IST-2022)* : матеріали конференції (м. Харків, 22–25 листопада, 2022 року). С. 23-27.
4. Ювченко К. С. Система розпізнавання емоцій людини з використанням алгоритмів глибокого навчання. *27-й Міжнародний молодіжний форум «Радіoeлектроніка та молодь у XXI столітті»* : зб. матеріалів форуму (м. Харків, 10–12 травня 2023 р.). Т. 7. Харків : ХНУРЕ, 2023. С. 235-236.
5. Ekman P. An argument for basic emotions. *Cognition and Emotion*. 1992. Vol. 6. P. 169-200.
6. Kim J., Lee D. Facial expression recognition robust to occlusion and to intra-similarity problem using relevant subsampling. *Sensors*. 2023. Vol. 23, № 5. 2619 p.
7. Ekman P., Friesen W., Hager J. The Facial Action Coding System: A Technique for the Measurement of Facial Movement. *Consulting Psychologists Press*. San Francisco, CA, USA. 2002. P. 313-352.
8. Facial Emotion Recognition for Photo and Video Surveillance Based on Machine Learning and Visual Analytics / Kalyta O.; Barmak O.; Radiuk P.; Krak I. *Applied Sciences*. 2023. Vol. 13, № 17. 9890 p.

9. Capturing violence in the night-time economy: A review of established and emerging methodologies / Philpot R., Liebst L., Møller K., Lindegaard M., Levine M. *Aggression and Violent Behavior*. 2019, Vol. 46, № 2. P. 56–65.

10. Bera A., Randhavane T., Manocha D. The emotionally intelligent robot: Improving socially-aware human prediction in crowded environments. *In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. CM3K-2019 (Long Beach, CA, USA, 16–20 June, 2019). P. 1–6.

11. UAV surveillance for violence detection and individual identification / Srivastava A., Badal T., Saxena P., Vidyarthi A., Singh, R. *Automated Software Engineering*. 2022, Vol. 29, № 1. 28 p.

12. Kazemi V., Sullivan J. One millisecond face alignment with an ensemble of regression trees. *In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus. OH, USA, 23–28 June, 2014. P. 1867–1874.

13. FACTS-IR: Fairness, accountability, confidentiality, transparency, and safety in information retrieval / Olteanu A., Garcia-Gathright J., de Rijke M., Ekstrand M.D., Roegiest A., Lipani A., Beutel A., Lucic A., Stoica A., Das A. *SIGIR Forum*. 2021. Vol. 53, № 2. P. 20–43.

14. Rangulov D., Fahim M. Emotion Recognition on Large Video Dataset Based on Convolutional Feature Extractor and Recurrent Neural Network. *Innopolis University*. 2020.

15. Introducing the recola multimodal corpus of remote collaborative and affective interactions / Ringeval F., Sonderegger A., Sauer J., Lalanne D. *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*. 2013. P. 1-8.

16. Recognition of affect in the wild using deep neural networks / Kollias D., Nicolaou M. A., Kotsia I., Zhao G., Zafeiriou S. *In Computer Vision and Pattern Recognition Workshops*. 2017. P. 1972–1979.

17. Aff-wild: Valence and arousal in-the-wild challenge / Zafeiriou S., Kollias D., Nicolaou M. A., Papaioannou A., Zhao G., Kotsia I. Aff-wild. *In Computer Vision and Pattern Recognition Workshops*. 2017. P. 1980–1987.

18. Khorrami P. How deep learning can help emotion recognition. PhD thesis, University of Illinois at Urbana-Champaign, 2017.
19. Attention is all you need / Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. *In Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 4–9 December, 2017. 11 p.
20. Subject independent facial expression recognition with robust face detection using a convolutional neural network / Matsugu M., Mori K., Mitari Y., Kaneda Y. *Neural Netw.* 2003. Vol. 16, № 5-6, P. 555–559.
21. Panagiotis A. Panagiotis F., Petros M. Exploiting Emotional Dependencies with Graph Convolutional Networks for Facial Expression Recognition. *16th IEEE International Conference on Automatic Face and Gesture Recognition*. ArXiv. 2021. P. 1-8.
22. Ryumina E., Dresvyanskiy D., Karpov A. In search of a robust facial expressions recognition model: A large-scale visual cross-corpus study. *Neurocomputing*. 2022. Vol. 514, № 4. P. 435-450.
23. Jersey Number Recognition with Semi-Supervised Spatial Transformer Network / Li G., Xu S., Liu X., Li L., Wang C. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Salt Lake City, UT, USA, 18–22 June, 2018. P. 690–696.
24. Minaee S., Minaei M., Abdolrashidi A. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. *Sensor*. 2021. Vol. 21, № 9. 3046 p.
25. Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition / Wang K., Peng X., Yang J., Meng D., Qiao Y. *IEEE Transformation Image Process*. 2020. Vol. 29, № 1, P. 4057–4069.
26. Wang C., Wang Z., Cui D. Facial Expression Recognition with Attention Mechanism. *In Proceedings of the 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*. CISP-BMEI (Shanghai, China, 23–25 October, 2021). P. 1–6.

27. Rojas R. *Neural Networks: A Systematic Introduction*. Springer. 1996. 61 p.
28. Mijwel M., Esen A., Shamil A. Overview of neural networks. *Computer Engineering Techniques Department*. 2019. P. 28-40.
29. Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU / Morteza Pour F., Perumal T., Mustapha N., Mohamed R. A. *University Putra Malaysia (UPM)*. 2023. 17 p.
30. Ke K., Huang M. Quality prediction for injection molding by using a multi-layer perceptron neural network. *Polymers*. 2020. Vol. 12, № 8. P. 1812-2020.
31. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects / Li Z., Liu F., Yang W., Peng S., Zhou J. *IEEE Trans Neural Network Learning System*. 2022. Vol. 33, № 12. P. 6999-7019.
32. Review of image classification algorithms based on convolutional neural networks / Chen L., Li S., Bai Q., Yang J., Jiang S., Miao Y. *Remote Sensing*. 2021. Vol. 13, № 22. 4712 p.
33. Dumoulin V., Visin F. A guide to convolution arithmetic for deep learning. *ArXiv*, 2016. 31 p.
34. Rainfall–runoff modelling using Long ShortTerm Memory (LSTM) networks / Kratzert F., Klotz D., Brenner C., Schulz K., Herrnegger M. *Hydrology and Earth System Sciences*. 2018 Vol. 22, № 11. P. 6005-6022.
35. Data Preprocessing: Definition, Key Steps and Concepts. URL: <https://www.techtarget.com/searchdatamanagement/definition/data-preprocessing> (дата звернення: 03.01.2024).
36. Data Preprocessing Techniques. Data preprocessing is a Data Mining. URL: <https://medium.com/almabetter/data-preprocessing-techniques-6ea145684812> (дата звернення: 03.01.2024).
37. Data Preprocessing in Machine Learning. URL: <https://www.v7labs.com/blog/data-preprocessing-guide> (дата звернення: 03.01.2024).

38. Regression Model Evaluation Metrics. URL: <https://medium.com/@brandon93.w/regression-model-evaluation-metrics-r-squared-adjusted-r-squared-mse-rmse-and-mae-24dcc0e4cbd3> (дата звернення: 03.01.2024).

39. Competition. Affective Behavior Analysis in-the-wild. URL: <https://ibug.doc.ic.ac.uk/resources/fg-2020-competition-affective-behavior-analysis/> (дата звернення: 07.01.2024).