Одеський національний університет імені І. І. Мечникова
Факультет математики, фізики та інформаційних технологій
Кафедра оптимального керування та економічної кібернетики

Кваліфікаційна робота

на здобуття ступеня вищої освіти «магістр»

«Задачі, моделі і методи обробки аудіосигналів»

«Problems, models and methods of processing audio signals»

Виконала: здобувачка денної форми навчання
спеціальності 113 Прикладна математика
Освітня програма «Прикладна математика»
Ларікова Валерія Володимирівна

Керівник: канд. тех. наук, доц. Мороз В. В.
Рецензент: канд. тех. наук, доц. Мазурок І. Є.

Рекомендовано до захисту:

Протокол засідання кафедри

№ ____ від _____ 2022 р.

Завідувач кафедри

_____ _____

Захищено на засіданні ЕК № _____

Протокол № ____ від _____ 2022 р.

Оцінка _____ / _____ / _____

Голова ЕК

_____ _____

Одеса — 2022 р.

ЗМІСТ

Вступ 4

1 Існуючі задачі класиікації звукових даних 6

1.1 Аналіз предметної області 6

1.2 Аналіз існуючих задач, параметрів та застосування аналізу звукових сигналів 10

1.3 Технологія DeepFake 15

1.4 Висновки до розділу 18

2 Методи, параметри та моделі класифікації аудіосигнлалів 20

2.1 Акустично-фонетичний підхід 21

2.2 Підхід розпізнавання образів 22

2.3 Підхід на основі шаблону 24

2.4 Статистичний підхід 25

2.5 Підхід штучного інтелекту (підхід на основі знань) 26

2.6 Підхід ANN 26

2.7 Підхід на основі HMM 27

2.7.1 Вилучення ознак 30

2.7.2 Узгодження ознак 31

2.7.3 Сполучення слово - фонема 32

2.8 Висновки до розділу 33

3 Задача виявлення та видалення імпульсного шуму в аудіосигналах 35

3.1 Задача виділення акустичного сигналу із суміші 35

3.1.1 Етап виявлення імпульсних шумів на основі авторегресивного моделювання 37

3.1.2 Оцінка параметрів авторегресії 40

3.2 Пряме порогове визначення 40

3.3 Висновки до розділу 43

4 Розробка програмноо забезпечення для виявлення та видалення імпульсних шумів 44

4.1 Проектування архітектури програмного забезпечення 44
4.2 Стаціонарне виявлення та видалення шумів 45
4.3 Нестаціонарне виявлення та видалення імпульсних шумів . . 47
4.4 Вибір технологій розробки 48
4.4.1 Мова програмування Python 49
4.4.2 Бібліотека Librosa 49
4.4.3 Бібліотека NumPy 50
4.5 Бібліотека Matplotlib 50
4.6 Бібліотека SciPy 51
4.7 Програмна реалізація застосунку для виявлення та видалення імпульсних шумів 52
4.8 Тестування розробленого застосунку 58
4.9 Порівняльний аналіз розроблених алгоритмів 63

Висновки 65

Список літератури 67

Додаток 70

ВСТУП

Актуальність теми дослідження. Протягом багатьох років вирішувалася проблема ізоляції джерел звуку та видалення корисного сигналу з акустичної суміші [1-3]. Здатність людського мозку відокремлювати призначений сигнал від різноманітних джерел перешкод у складному гучному слуховому середовищі демонструє феноменальний результат. Однак, це складна проблема в області цифрової обробки сигналів. Значна частина навколишнього штучного акустичного середовища, яка оточує людину, створюється в результаті широкого розповсюдження звукотехнічної техніки в поєднанні із засобами масової аудіовізуальної інформації та комунікації. Таке середовище, як правило, позитивно впливає як емоційно, так і раціонально.

З розповсюдженням мультимедійних можливостей персональних комп'ютерів цифрова обробка аудіосигналів стає все більш важливою роллю. Коли в середині 20 століття були винайдені перші комп'ютери, були закладені основи цифрової обробки. Тепер, коли навіть настільні комп'ютери мають обчислювальну здатність виконувати складну обробку сигналів у реальному часі, питання якості алгоритму, а не швидкості алгоритму є важливішим. Алгоритми цифрової обробки переходять від вузькоспеціалізованих до широко використовуваних. Раніше вони використовувалися для радарів, професійного звукозапису та обробки звуку. У сучасному мультимедійному інтерфейсі „людина-машина", який включає мікросхеми цифрової камери, що обробляють зображення, мобільні телефони, які кодують і обробляють звук, і персональні комп'ютери, які можуть функціонувати як центр домашніх розваг завдяки своїм широким можливостям обробки звуку, зображень і відео, алгоритми цифрової обробки звукових даних використовуються все частіше.

У кваліфікаційній роботі розглянута задача виявлення та видалення імпульсного шуму в аудіосигналах з використанням методів авторегресії. На основі проведеного дослідження розроблено власне програмне рішення, яке дозволяє накладати імпульсний шум на звукові сигнали, а також виявляти та видаляти локальні деградації для покращення якості сприйняття запису.

Враховуючи вищевикладене, зроблено висновок, що тема дипломної роботи „Задачі, моделі і методи обробки аудіо сигналів” є актуальною на момент написання даної роботи.

Мета роботи: Дипломна робота спрямована на дослідження існуючих систем і методів обробки акустичних сигналів з використанням технологій машинного навчання, а також на розробку нових або покращення функціональності існуючих методів обробки акустичних сигналів із застосуванням програмного та апаратного підходу, що дозволяє точніше розпізнавати звукові дані.

Об’єкт дослідження: Виявлення та видалення імпульсного шуму в аудіосигналах за допомогою технологій машинного навчання. Предмет дослідження: Методи і алгоритми машинного навчання для виявлення та видалення імпульсного шуму в аудіосигналах різноматного природного походження.

У 2022 році дана робота пройшла апробацію на двадцять другій міжнародній науково-технічній конференції, яка проводилася на базі національного технічного університету «Харківський політехнічний інститут», та має назву «Проблеми інформатики та моделювання». Тези були опубліковані у збірнику [27]. За результатами дослідження, яке було проведено у даній роботі, було написано статтю, яка має назву «Видалення імпульсних шумів в аудіосигналах» . Стаття була опублікована у збірнику [28].

ВИСНОВКИ

Отже, на основі дослідження проведеного у кваліфікаційній роботі, можна зробити наступні висновки:

1) В першому розділі було проведено аналіз предметної області, а також розглянуті основні задачі, параметри та перспективи застосування аналізу звукових даних з використанням технологій машинного навчання. Проаналізована технологія DeepFake, як один з найпопулярніших в сучасності інструментів для підробки голосів різніх людей. Визначено, що протягом багатьох років багато дослідників продемонстрували, що підходи до машинного навчання точніші, ніж методи глибокого навчання, незалежно від використовуваних функцій. Однак масштабованість методів машинного навчання не підтверджена через надмірне навчання та ручне вилучення функцій, особливо з великою кількістю аудіофайлів.

2) В другому розділі розглянуті основні методи та моделі класифікації аудіосигналів. У цьому розділі описано кілька методів розпізнавання голосу, включаючи багато фундаментальних підходів до розпізнавання мовлення: акустично-фонетичний метод; метод розпізнавання образів; метод штучного інтелекту; метод нейронної мережі; метод прихованої моделі Маркова (HMM).

Результати аналізу демонструють, що робота зі створення систем на основі технологій глибинного навчання для обробки звукових сигналів, які б покращили існуючі системи розпізнавання мовлення в цілому, зараз є актуальною. Одним з рішень для подолання цієї проблеми є застосування методів шумозаглушення на основі методів глибинного навчання для побудови методу обробки звукових сигналів, що покращить точність розпізнавання голосових сигналів поточними системами.

3) В третьому розділі було розглянуто та деталізовано архітектуру одного з відомих методів виявлення та видалення імпульсних шумів. Алгоритм, представлений у цьому розділі, є простим, але ефективним методом відновлення звукових сигналів, пошкоджених імпульсним шумом. Обмежений діапазон параметрів і той факт, що фази виявлення та інтерполяції мають спільні параметри, дозволяють швидко обробляти сигнали в авто-

матичному режимі. Результати показують, що метод здатний працювати з різними типами сигналів (класика, джаз, вокал) і різними типами погіршень.

Метод, описаний у цьому розділі, можна використовувати для нормалізації аудіосигналів, які були пошкоджені імпульсним шумом, наприклад клацанням, сплеском або подряпинами.

4) В четвертому розділі проведено моделювання архітектури основних алгоритмів програмного застосунку з виявлення та видалення імпульсних шумів, охарактеризовано основні технології, які були використані при розробці, а також представлені приклади обробки короткого та довгого аудіо.

На основі зазначено в четверому розділі аналізу були зроблені наступні висновки відносно алгоритмів:

1) Стаціонарні та нестаціонарні алгоритми можна використовувати при обробці аудіофайлів будь-якої довжини.
2) Якщо для користувача важлива швидкість обробки даних, то необхідно вибрати нестаціонарне виявлення та видалення імпульсних шумів.
3) При використанні нестаціонарного виявлення та видалення імпульсних шумів відсутня вирогідність появи ‘артефактів’ при обробці аудіо, а при використанні стаціонарного алгоритму така вирогідність присутня.

Для покращення роботи застосунку, в майбутньому, необхідно:

1) Додати можливість мультипроцесинга, для того, щоб у користувача була можливість працювати з ще більшим об’ємом даних.
2) Реалізувати GUI інтерфейс з інтерактивними елементами для зручності використання в повсякденному житті та при дослідженнях.
3) Розробити API інтерфейс для зручної роботи з ендпоінтами та підвищення швидкості виконання запитів.

СПИСОК ЛІТЕРАТУРИ

1. T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann. „Making machines understand us in reverberant rooms: robustness against reverberation for automatic speech recognition”, *IEEE Signal Processing Magazine,* vol. 29, no. 6, pp. 114–126, 2012.
2. Juin-Hwey Chen and Allen Gersho, „Adaptive postfiltering for quality enhancement of coded speech”, *IEEE Transactions on Speech and Audio Processing,* vol. 3, no. 1, pp. 59–71, 1995.
3. K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, „On the properties of neural machine translation: Encoder-decoder approaches”, *Proc. Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8),* 2014.
4. Р. Тейлор, *Шум.* Світ, 1978, ст. 308
5. В. В. Шелемаха, „Огляд класифікації шумів”, `https://ela.kpi.ua/bitstream/123456789/46751/1/P.76-78.pdf`, ст. 76 - 78.
6. What Is Brown Noise? Live Science [Online]. `https://www.livescience.com/38547-what-is-brown-noise.html`
7. What is Gray Noise? Techopedia [Online]. `https://www.techopedia.com/definition/27898/gray-noise`
8. Graves A., Mohamed A., Hinton G. „Speech recognition with deep recurrent neural networks”. `https://arxiv.org/pdf/1303.5778.pdf`
9. Olah C. „Understanding LSTM networks”. `https://colah.github.io/posts/2015-08-understanding-lstms/`
10. Karpathy A. „Understanding the unreasonable effectiveness of recurrent neural networks”. `http://karpathy.github.io/2015/05/21/rnn-effectiveness`
11. S. Boll, „Suppression of acoustic noise in speech using spectral subtraction”, *IEEE Transactions on acoustics, speech, and signal processing,* vol. 27, no. 2, pp. 113–120, 1979.
12. Suryo Wijoyo, „Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot”, *Internati-*

onal Conference on Information and Electronics Engineering IPCSIT, vol.6, IACSIT Press, pp.179-183, Singapore, 2011.

13. G. Gnaneswari, S. R. VijayaRaghava, A. K. Thushar, Dr.S.Balaji, „Recent Trends in Application of Neural Networks to Speech Recognition", *International Journal on Recent and Innovation Trends in Computing and Communication,* vol. 4. Issue 1, pp 18 - 25
14. Palaz, D., Doss, M., Collobert, R. „Convolutional Neural Networks-based continuous speech recognition using raw speech signal", *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* pp. 4295 - 4299
15. Abdel-Hamid O., Abdel-Rahman M., Hui J., Li D., Penn G., Yu D. „Convolutional neural networks for speech recognition", *IEEE/ACM Transactions on Audio, Speech and Language Processing,* vol. 22. Issue 10, October 2014 pp. 1533–1545
16. Abdel-Hamid O., Abdel-Rahman M., Hui J., Li D., Penn G., Yu D. „Convolutional neural networks for speech recognition", *IEEE/ACM Transactions on Audio, Speech and Language Processing,* vol. 22. Issue 10, October 2014, pp. 1533–1545
17. Laurent Oudre, „Automatic Detection and Removal of Impulsive Noise in Audio Signals", *Image Processing On Line,* vol. 5, (2015), pp. 267–281. `https://doi.org/10.5201/ipol.2015.64`
18. T. Pham and L. Sim, „Acoustic detection and tracking of small, low-flying threat aircraft", in Proc. ASC, Jan. 2005, pp. 1–2.
19. G. Secinti, P. B. Darian, B. Canberk, and K. R. Chowdhury, „SDNs in the sky: Robust end-to-end connectivity for aerial vehicular networks", *IEEE Wireless Commun. Mag.,* vol. 56, no. 1, pp. 16–21, Jan. 2018.
20. M. Peacock and M. N. Johnstone, „Towards detection and control of civilian unmanned aerial vehicles", *Proc. IWAR,* Dec. 2013, pp. 9–15.
21. T. Pham and N. Srour, „TTCP AG-6: Acoustic detection and tracking of UAVs", *Proc. SPIE 5417*, Sept. 2004, pp. 24–30.
22. Z. Kaleem and M. Rehmani, „Amateur drone monitoring: State-of-the-art architectures, key enabling technologies, and future research directions, *IEEE Wireless Commun.,* vol. 25, no. 2, pp. 150–159, May 2018.

23. V. K. Kakar and M. Kandpal, „Techniques of acoustic feature extraction for detection and classification of ground vehicles", *Int. J. Emerging Technology Advance Eng.,* vol. 3, no. 2, pp. 419–426, Feb. 2013
24. I. Patel and Y. S. Rao, „Speech recognition using hidden Markov model with MFCC-subband technique, *Proc. ITC,* Mar. 2010, pp.168–172.
25. Laurent Oudre, Automatic Detection and Removal of Impulsive Noise in Audio Signals, *Image Processing On Line,* vol. 5, Nov. 2015, pp. 267–281. `https://doi.org/10.5201/ipol.2015.64`
26. J. Hunt, *Advanced Guide to Python 3 Programming (Undergraduate Topics in Computer Science),* Firts Edition, 2019
27. НТУ «Харківський політехнічний інститут», „Проблеми інформатики та моделювання," Двадцять друга міжнародна науково-технічна конференція, Харків, 2022, с. 56. `https://drive.google.com/drive/folders/1e2rFD0PjysLpWKLID6UR0SU8P5PuaEHz?usp=sharing`