

ДОСЛІДЖЕННЯ КАПСУЛЬНИХ НЕЙРОМЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ ПРИРОДНОГО МОВЛЕННЯ

Петришин А.Ю.

Науковий керівник – канд. техн. наук, доц. Єсілевський В.С.
Харківський національний університет радіоелектроніки, каф. ПМ,
м. Харків, Україна
e-mail: andrii.petryshyn@nure.ua

This thesis focuses on evaluating neural network-based models for speech recognition under noisy conditions. It examines the effectiveness of the capsule neural network across various types and intensities of noise. By gathering and preprocessing a dataset of noisy speech recordings, the research compares these models based on their recognition accuracy, resilience to different noise types and intensities, and computational efficiency.

Сьогодні у суспільстві, штучний інтелект (ШІ) прогресує стрімкими темпами. Серед напрямків застосуванні ШІ є розпізнавання природного мовлення. Застосування такого ШІ може полегшити роботу інших ШІ: генеративні статистичні моделі із застосуванням технологій ШІ для створення нового тексту, зображень тощо. Серед інших прикладів є військове застосування в зонах, де рівень шуму великий і де простіше було б застосувати текстову форму замість вербальної: гелікоптер, промислові виробництва тощо.

Шум є ключовою проблемою у розпізнаванні мовлення, оскільки він може походити з різних джерел, таких як білий шум під час трансмісії сигналу чи фонові звуки, які складно відфільтрувати. Тому розробка методів, які можуть ефективно справлятися з цими видами шумів, є важливим напрямком дослідження.

У дослідженні розглядається розпізнавання природного мовлення із шумом методом капсульних нейронних мереж [1]. Така мережа має структуру наближену до людського ока на відміну від згорткових нейромереж, що зараз широко застосовуються.

Для безпосереднього аналізу моделі, у дослідженні збирають аудіозаписи з шумом [2], перетворюють ці аудіозаписи у зрозумілі для моделі дані і навчають модель для подальшого їх аналізу. Порівняння моделей відбувається за такими параметрами: точність розпізнавання, стійкість до шуму, ефективність обчислень, порівняння результатів на аудіозаписах із іншим шумом і рівнем шуму [3–4].

Обробка аудіозаписів відбувається шляхом перетворення їх у аудіосигнали, вилучення ознак: коефіцієнтів мел-частотного кепстру [5] (рис. 1). Готові дані використовують вже як вхідні значення на моделі нейронної мережі, які навчається на цих даних і далі використовується щоб розпізнавати природне мовлення.

В ході дослідження було зроблено висновок, що модель показала найкращі результати при рівні білого шуму 0.2 (де 1 – повне зашумлення). При збільшенні рівні шуму, якість моделі зменшується. На рівні шуму 0.3 модель показала себе найгіршою. При тому точність розпізнавання моделі не перевищила 80 % в цілому.

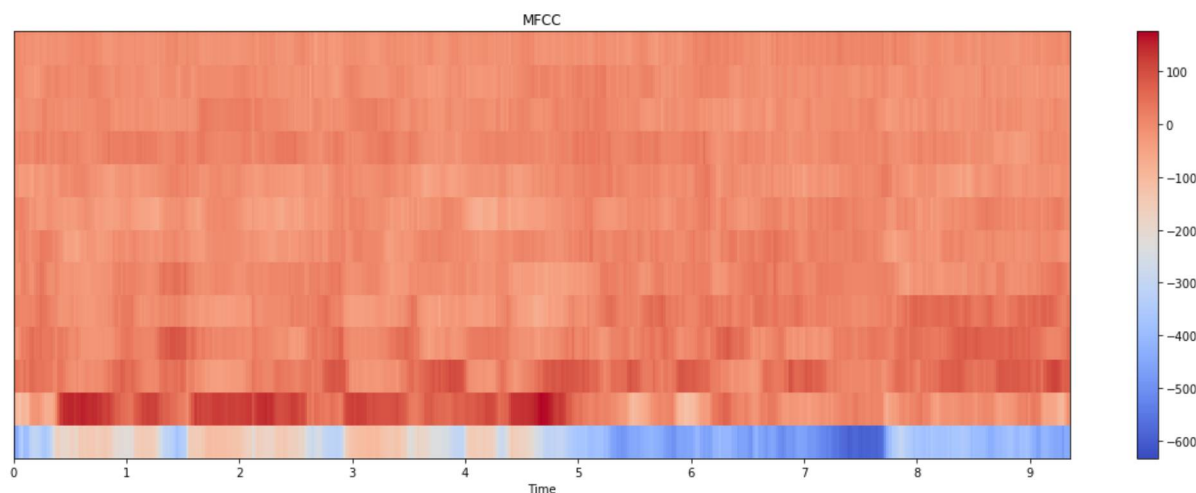


Рисунок 1 – Приклад коефіцієнтів мел-частотного кепстру

В даній роботі пропонується порівняльні результати роботи капсульної нейронної мережі на білому та рожевому шумах із різними рівнями шуму при розпізнаванні природного мовлення, що може далі бути застосовано для більш поглибленого вивчення даної нейромережі та дослідження проблеми шуму у природному мовленні.

Список використаних джерел:

1. Kurama V. Introduction to Capsule Networks | Paperspace Blog. Paperspace Blog. URL: <https://blog.paperspace.com/capsule-networks/> (дата звернення: 04.03.2024).

2. To make a better sensor, just add noise. ScienceDaily. URL: <https://www.sciencedaily.com/releases/2020/09/200903162041.htm> (дата звернення: 04.03.2024).

3. A simple model for pink noise from amplitude modulations. arXiv.org. URL: <https://arxiv.org/abs/2301.11176> (дата звернення: 04.03.2024).

4. Yuvchenko, K., Yesilevskiy, V., Sereda, O. (2022), "Human emotion recognition system using deep learning algorithms", Innovative Technologies and Scientific Solutions for Industries, No. 3 (21), P. 60–69. DOI:<https://doi.org/10.30837/ITSSI.2022.21.060>

5. Mel Frequency Cepstral Coefficient and its Applications: A Review. IEEE Xplore. URL: <https://ieeexplore.ieee.org/document/9955539> (дата звернення: 04.03.2024).