УДК 004.934:004.032.26

МАТЕМАТИЧНІ МОДЕЛІ ТА МЕТОДИ РОЗПІЗНАВАННЯ ПРИРОДНОГО МОВЛЕННЯ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ

Петришин А.Ю.
Науковий керівник – канд. техн. наук, доц. Єсілевський В.С.
Харківський національний університет радіоелектроніки, каф. ПМ,
м. Харків, Україна
тел. +380938025430, email: andrii.petryshyn@nure.ua

This thesis centers around the comparison of neural network-based models for speech recognition in the presence of noise. The study reviews denoising autoencoders, CNNs, RNNs, and transformer-based architectures, and evaluates their suitability for different noise types and levels. Comparison is conducted by collecting and preprocessing dataset of noisy speech recordings and experimenting to compare the performance of these models in terms of recognition accuracy, robustness to different noise types and levels, and computational efficiency.

В сучасному суспільстві розпізнавання живого мовлення є актуальною темою наприклад для збереження телефонних розмов у текстовій формі, загального транскрибування аудіозаписів з подальшою їх роботою (переклад текстів тощо). Подальші можливості роботи з розпізнаним текстом необмежені: створення асистента, збереження основних інформації записів у текстовій формі для зручного пошуку тощо.

Під час розпізнавання живого мовлення виникає проблема шуму, який утворюється з різних причин: передача сигналу утворює білий шум; природні умови утворюють натуральний шум у вигляді вітру, водоспаду та іншого, розмови людей на фоні запису теж вважається шумом.

У дослідженні розглядається розпізнавання природного мовлення з шумом методами на основі нейронних мереж [1]. Серед цих методів глибинне нейронні мережі [2], рекурентні нейронні мережі [3], згорткові нейронні мережі [4] і нейронні мережі із застосуванням трансформаторів.

Для безпосереднього порівняння різних моделей, у дослідженні збирають аудіозаписи з шумом, обробляють їх у зрозумілі для нейронних мереж дані і навчають моделі для подальшого їх аналізу. Порівняння моделей відбувається за такими параметрами: точність розпізнавання, стійкість до шуму, ефективність обчислень.

Обробка аудіозаписів відбувається шляхом перетворення їх у аудіосигнали, вилучення ознак, які можна представити у числовій формі, таких як, спектрограми [5] (рис. 1) тощо. Готові дані використовують вже як вхідні значення на моделі нейронних мереж, які навчаються на цих даних і далі використовуються щоб розпізнавати мовлення.

В ході дослідження було зроблено висновок, що модель глибинних нейронних мереж показує себе найгірше в порівнянні з рекурентними і

згортковими нейронними мережами і моделі з трансформаторами. Найкращою моделлю виявилась модель із застосування перетворювача (трансформатора). Точність жодної з моделей не перевищила 80 %, що показує, що для практичного застосування в більшості випадків вони не є зручними. Такі моделі можна застосовувати коли треба розпізнати основну частину тексту для загального розуміння про що йде мова в аудіозаписі.

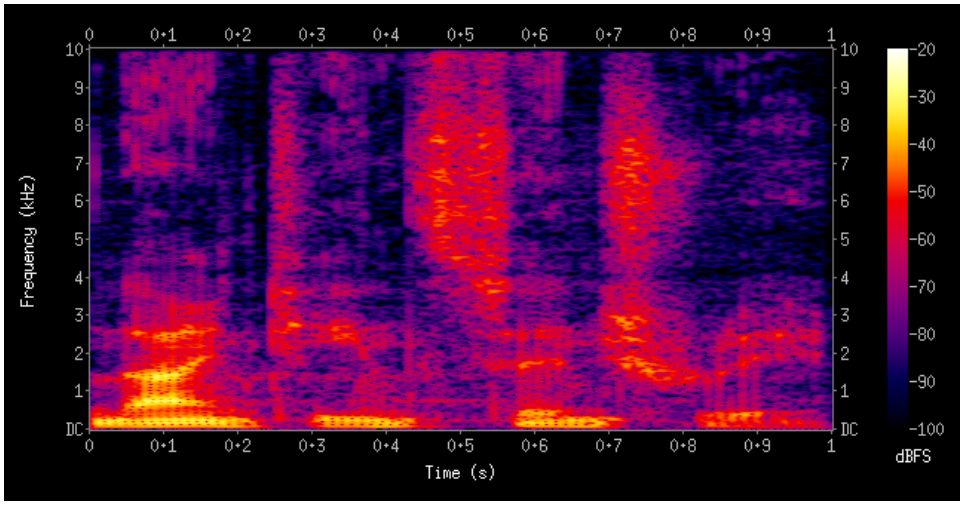

Рисунок 1 – Приклад спектрограми

Порівняння моделей на основі нейронних мереж для розпізнавання живого мовлення покаже їхні сильні і слабкі сторони для розпізнавання мовлення з шумом і практичність їх застосування.

Список використаних джерел:

1. Hannun, A., CASE, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., & Coates, A. (2014). *Deep speech: Scaling up end-to-end speech recognition*.

2. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., & Sainath, T. N. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine*.

3. *Towards data science* (2018, 27 December). Recognizing Speech Commands Using Recurrent Neural Networks with Attention https://towardsdatascience.com/recognizing-speech-commands-using-recurrent-neural-networks-with-attention-c2b2ba17c837

4. Collobert, R., Puhrsch, C., & Synnaeve, G. (2016, 13 September). *Wav2Letter: an End-to-End ConvNet-based Speech Recognition System*.

5. Towards data science (2020, 19 July). *Understanding Audio data, Fourier Transform, FFT and Spectrogram features for a Speech Recognition System*. https://towardsdatascience.com/understanding-audio-data-fourier-transform-fft-spectrogram-and-speech-recognition-a4072d228520