

РОЗПІЗНАВАННЯ МОВЦЯ ЗА АУДІОЗАПИСОМ З ВИКОРИСТАННЯМ КЕПСТРАЛЬНИХ КОЕФІЦІЄНТІВ ТА ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ

Матиченко А.Д., Полякова М.В.

Національний університет «Одеська політехніка», м. Одеса

Розглянуто питання структурного налаштування згорткової нейронної мережі з [1] для розпізнавання мовців у просторі мелчастотних кепстральних коефіцієнтів. Метою роботи було зменшення кількості параметрів нейронної мережі і, як наслідок, скорочення часу обробки аудіозапису за умови достатньої якості розпізнавання (процент правильного розпізнавання вище за 95%).

Запропонована у результаті структурного налаштування нейронна мережа має менше шарів, ніж архітектура нейронної мережі з [1]. Замість функції активації ReLU застосовано споріднену до цієї функції функцію Leaky ReLU з параметром 0.1 (рис. 1). Змінено кількість фільтрів та розмірність ядер в згорткових шарах. Збільшено розмірність ядер для пулінгового шару з обранням максимального елементу. Запропоновано використання усереднення результатів кожної згортки для переходу від двовимірної згортки до повнозв'язного шару з функцією активації Softmax.

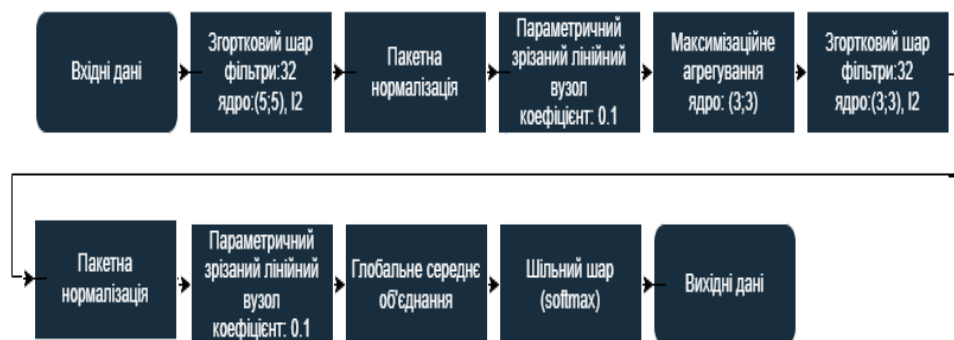


Рисунок 1 – Архітектура запропонованої згорткової нейронної мережі

Під час експерименту обчислювалася кількість параметрів базової та запропонованої нейронної мережі. Запропонована нейронна мережа потребує налаштування 4268 параметрів. Нейронна мережа з [1] потребує налаштування 6000 параметрів. Таким чином, кількість параметрів запропонованої нейронної мережі менша на 29 відсотків кількості параметрів базової нейронної мережі за умови майже однакової якості розпізнавання мовців.

Окрім того, на п'яти датасетах аудіозаписів, що відповідали 2-м, 2-м, 5-ти, 20-ти, 29-ти дикторам, оцінювався час навчання запропонованої та базової нейронної мережі. Було отримано скорочення часу навчання запропонованою мережею на 17% у порівнянні з мережею з роботи [1].

Література:

1. Bunrit S., Inkian T., Kerdprasop N., Kerdprasop K. Text-independent speaker identification using deep learning model of convolution neural network. *International Journal of Machine Learning and Computing*. 2019. Vol. 9, No. 143-148. DOI: 10.18178/ijmlc.2019.9.2.778