

ДОСЛІДЖЕННЯ МЕТОДІВ ТА МОДЕЛЕЙ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ УЧАСНИКІВ АУДІО РОЗМОВИ

Андрєєв І. Г.

Науковий керівник – к.т.н., доц. каф. ПІ Каук В. І.

Харківський національний університет радіоелектроніки, каф. ПІ,
м. Харків, Україна

e-mail: ivan.andrieiev@nure.ua

The object of the study is the process of automatic speech recognition (ASR) and one of the branch of it – speaker recognition or diarization (SD).

The purpose of the work is to study the stages of the process of speech and speakers recognition, the analysis of methods and models of machine learning and neural networks, as well as modern frameworks for training speech recognition models and diarization. Based on the received results of the study it would be selected the existing models, systems or products that would play a role of starting point in the creation of a custom ASR and SD models for enhancing those processes and bringing more values and benefits in different areas of human being.

У сучасному цифровому світі, де аудіо розмови і мовлення є невід'ємною частиною нашого повсякденного життя, проблема розпізнавання мовлення учасників розмови стає все більш актуальною і важливою. Ця проблема має великий потенціал застосування в різних сферах, включаючи телефонну комунікацію, відеоконференції, медіа і багато інших галузей.

Машинне навчання та нейронні мережі є ключовими компонентами сучасних процесів розпізнавання мовлення та визначення учасників розмови. Машинне навчання дозволяє моделі навчатися на основі великої кількості даних та робити прогнози, а нейронні мережі, зокрема глибокі нейронні мережі, є потужними інструментами для обробки аудіо даних та розпізнавання мовлення.

Покращення процесу розпізнавання учасників розмови та оптимізація використання цих технологій можуть дати позитивний вплив у таких напрямках:

- аутентифікація та безпека: питання безпеки стає особливо актуальним у контексті конфіденційності та захисту від несанкціонованого доступу;
- навчання: транскрипція запитань і відповідей студента для аналізу відповідей, наданих професором або студентами;
- здоров'я: окремі коментарі пацієнтів і лікарів як для особистих зустрічей, так і для телефонних консультацій;

- підтримка продажів: відстеження того, хто що сказав на зустрічі з продавцями, і навчання продавця, що говорити та коли мовчати;
- аналіз доповідача: відстежуйте поточні та попередні коментарі певного доповідача під час зустрічей або відстежуйте час розмови під час телефонної розмови, тощо.

Метою цього дослідження є аналіз методів та моделей автоматичного розпізнавання учасників аудіо розмови з використанням машинного навчання і нейронних мереж та встановленню шляхів покращення окремих етапів або всього процесу розпізнавання мовлення загалом. До основних підходів визначення мовлення виділені:

- підхід під наглядом (класифікація): модель навчена розпізнавати обмежену кількість промовців (тобто класів). Цей підхід менш гнучкий, але може бути дуже ефективним, особливо для великої кількості промовців (>5);

- підхід без нагляду (кластеризація): модель групує аудіо сегменти відповідно до промовця на основі виділених аудіо характеристик.

Об'єктом дослідження є процеси автоматичного розпізнавання мовлення (ASR) та розпізнавання учасників розмови (SD), а також існуючі системи та сервіси, що реалізують ці процеси. Найбільш популярні сервіси з розпізнавання мови та визначення учасників розмови базують навчання своїх моделей на основі згорткових (CNN) і рекурентних типів нейронних мережах (RNN).

Отримані результати цього дослідження стануть потенційно важливою складовою для розвитку технологій обробки аудіо даних та розпізнавання мовлення. Аналіз моделей та процесів, з одного боку, є важливою складовою на шляху до повного розуміння цієї галузі та її перспектив, а з іншого, є реальним шансом використати знання з області машинного навчання та нейронних мереж для вирішення реальних завдань сучасного світу аудіо відносин.

Список використаних джерел

1. Бريدінський, В. Побудова системи ідентифікації мовців на основі бібліотеки аудіообробки PyAnnote. Information Technology: Computer Science, Software Engineering and Cyber Security. 2022. №2. С. 3-11.
2. Корнієнко О. Метод відображення мовних сигналів у задачі розпізнавання мовця. Технічні науки та технології. 2017. №3. С. 129-137.
3. Bai Z., Zhang X.-L. Speaker recognition based on deep learning: An overview. Neural Networks. 2021. Vol. 140. P. 65-99.
4. Introduction to Speech Processing. Speaker Diarization. URL: <https://speechprocessingbook.aalto.fi> (дата звернення: 20.11.2023).