Д. О. Кузнєцов

Науковий керівник - д.т.н., професор Почапський Є. П.

Програмна система створення субтитрів із використанням моделей машинного навчання

Стан проблеми. Голосовий переклад є перспективною технологією, яка дозволяє перекладати мовлення з однієї мови на іншу в режимі реального часу. Розвиток штучного інтелекту та машинного навчання значно покращив можливості таких систем [1]. Сучасні системи голосового перекладу мають широкий спектр застосувань, включаючи міжнародне спілкування, бізнес-комунікації, освіту, подорожі.

Машинне навчання відіграє ключову роль у розвитку систем голосового перекладу, дозволяючи їм адаптуватися до різноманітних мовних особливостей, виявляти контекстуальні зв'язки в мовленні та покращувати точність перекладу [2]. Однак, реалізація швидкої та точної системи розпізнавання голосу залишається складним завданням при розробці програм голосового перекладу [3].

Існує ряд популярних систем розпізнавання мовлення на базі штучного інтелекту, таких як Google Cloud Speech-to-Text, Wit.ai, Microsoft Azure AI Speech, OpenAI Whisper та інші. Використання готових моделей розпізнавання мовлення дозволяє уникнути складного та часозатратного процесу створення та навчання власних моделей [4].

Постановка задачі. Розробити архітектуру програмної системи створення субтитрів із використанням моделей машинного навчання. Система повинна вміти записувати звуковий потік з будь-якого додатку, розпізнавати в ньому мовлення, генерувати текстове представлення, перекладати його та виводити на екран. Важливою вимогою є робота системи в режимі реального часу.

Розв'язання задачі. Розроблено структурну схему програми, що складається з чотирьох модулів:

1. Основний модуль для взаємодії з користувачем та керування потоками виконання інших модулів.
2. Модуль захоплення звуку для запису аудіоданих з системного звукового потоку.
3. Модуль розпізнавання мовлення на базі штучного інтелекту для перетворення аудіо в текст.
4. Модуль перекладу тексту для забезпечення міжмовного перекладу.

Програма використовує паралельні потоки виконання та

спільні ресурси (контейнери) для обміну даними між модулями, що забезпечує роботу в режимі реального часу. Розроблено три типи контейнерів:

- Контейнер аудіоданих: масив байтів для зберігання записаного звуку.
- Контейнер розпізнаного тексту: змінна рядкового типу для зберігання розпізнаного тексту.
- Контейнер перекладеного тексту: для зберігання перекладеного тексту перед виводом на екран.

Алгоритм роботи програми включає:

1. Запуск основного потоку та очікування активації перекладу користувачем.
2. Паралельний запуск потоків захоплення звуку, розпізнавання мовлення та перекладу.
3. Постійне оновлення контейнерів даними: запис звуку, розпізнавання мовлення, переклад тексту.
4. Виведення перекладеного тексту на екран основним модулем.
5. Очищення контейнерів після завершення кожного речення (визначається наявністю крапки в кінці тексту).

Особливістю розробленої системи є її здатність працювати з потоковими даними, що дозволяє обробляти мовлення в реальному часі без необхідності очікування завершення всього аудіозапису.

Висновки. Запропонована архітектура програмної системи дозволяє ефективно створювати субтитри в режимі реального часу, використовуючи моделі машинного навчання для розпізнавання мовлення та перекладу. Система демонструє потенціал для застосування в різних сферах, де потрібен миттєвий переклад аудіоконтенту, включаючи міжнародні конференції, освітні платформи, та системи синхронного перекладу для медіаконтенту.

Література.

1. Cheng, Q., et al. (2019). Breaking the Data Barrier: Towards Robust Speech Translation via Adversarial Stability Training.
2. Sperber, M., et al. (2017). Neural lattice-to-sequence models for uncertain inputs.
3. Nakamura, S., et al. (2005). The ATR Multilingual Speech-to-Speech Translation System.
4. Xu, C., et al. (2023). Recent Advances in Direct Speech-to-text Translation.
5. Miner, A. S., et al. (2020). Assessing the accuracy of automatic speech recognition for psychotherapy.