

АНАЛИЗ СОВРЕМЕННЫХ ПРОГРАММНЫХ СРЕДСТВ ДЛЯ РЕШЕНИЯ ЗАДАЧ РАСПОЗНАВАНИЯ РЕЧИ

В статье рассматриваются современные программные средства для решения задач распознавания речи с учётом современных подходов к работе с речевыми потоками, обработки и выделения основных признаков из аудио сигналов.

ВВЕДЕНИЕ

Распознавание речи или автоматическое распознавание речи (ASR) играет важную роль в реализации интерфейса человек-компьютер. Технология распознавания речи уже используется в большом количестве приложений. Распознавание речи используется для формирования субтитров, в голосовых помощниках и робототехнике.

I. СОВРЕМЕННЫЕ АРХИТЕКТУРЫ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ РЕЧИ

Как видно из общей схемы системы распознавания речи представленной на рис.1, работа подобных систем состоит из двух фаз, называемых предварительной обработкой и пост-обработкой. Первая часть системы извлекает характеристики речевого сигнала, а вторая - для строит речевую модель и находит наиболее вероятный вариант представления речи в виде текста. Как правило часть системы, которая занимается непосредственно распознаванием состоит из акустической модели, языковой модели и словаря.



Рис. 1 – Архитектура ASR системы

Таким образом построение ASR системы можно свести к выбору соответствующего метода извлечения значимых характеристик речевого потока и построения части ответственной за распознавание.

Можно выделить следующие методы распознавания:

- метод основанный на шаблонах;
- статистические методы;
- скрытые марковские модели (HMM);
- нейронные сети.

Лежнев Антон Викторович, магистрант кафедры информационных технологий автоматизированных систем БГУИР, anton.lezhnev@yandex.by.

Научный руководитель: Захарьев Вадим Анатольевич, кандидат технических наук, доцент кафедры систем управления БГУИР, zahariev@bsuir.by.

В последние годы с развитием технологий наибольшее развитие получили способы построения ASR систем с использованием нейронных сетей, подходы основанные на нейронных сетях позволяют работать с зашумлённым речевым потоком в различных условиях.

Далее будут рассмотрены программные средства по распознаванию речи.

II. ПРОГРАММНЫЕ ИНСТРУМЕНТЫ ДЛЯ СИСТЕМ РАСПОЗНАВАНИЕМ РЕЧИ

Среди большинства библиотек для работы с нейросетями можно выделить tensorflow, pytorch. С помощью библиотеки pytorch можно реализовать несколько моделей для решения задачи ASR, в частности CTC и seq2seq. Для сравнения эффективности работы различных моделей и их программных реализаций сравним точность распознавания нескольких моделей.

Таблица 1 – Результаты работы моделей

Система	LER
HMM	38.85%
BLSTM/HMM	35.21%
CTC	30.51%

Как видно исходя из результатов работы различных моделей на наборе данных TIMIT, представленные выше модели справляются с работой лучше, чем классические алгоритмы.

ЗАКЛЮЧЕНИЕ

Предлагаемый набор программных инструментов и подходов описанных выше, позволяет решить большинство современных задач по распознаванию речи и смежных задач по работе с речевым потоком.

Список литературы

1. Speech Recognition Challenges by using Neural Network Approaches – Modeof access: <https://www.researchgate.net/publication/331275489>
Date of access: 20.01.2020
2. Comparative Study on Feature Extraction Techniques in Speech Recognition – Modeof access: <https://www.researchgate.net/publication/278549945>
Date of access: 20.01.2021