Бауманский учебный центр «Специалист»

Программа курса Нейронные сети. Уровень 3. Компьютерная обработка звука

Код курса: **ПИМЛ4**

Тема	Ак. часов
Модуль 1. Введение в обработку звука • Основы обработки звуковых данных. Природа звука: частота, амплитуда, временные ряды. Преобразование сигналов: STFT, спектрограмма, мел-спектрограмма. • **Лабораторная работа:** преобразование звуковых файлов в спектрограммы и их визуализация (наборы данных ESC-50 Dataset - звуки окружающей среды, UrbanSound8K Dataset - городские звуки). • Работа с аудиоданными в Python. Библиотеки для обработки звука: Librosa, torchaudio. Методы предобработки: нормализация, фильтрация шума. • **Лабораторная работа:** предобработка аудиоданных из общедоступного набора данных (наборы данных ESC-50 Dataset - звуки окружающей среды, UrbanSound8K Dataset - городские звуки).	**4**
Модуль 2. Базовые нейронные сети для обработки звука • Рекуррентные нейронные сети (RNN) в задачах обработки звука. Применение RNN, LSTM, GRU для временных рядов. Задачи классификации звуков. • **Лабораторная работа:** реализация модели LSTM для классификации звуков на основе спектрограмм. • Свёрточные нейронные сети (CNN) для обработки звука. Использование CNN для анализа спектрограмм. Преимущества CNN в задачах локальных зависимостей. • **Лабораторная работа:** обучение свёрточной нейронной сети для классификации команд из набора Google Speech Commands.	**4**
Модуль 3. Современные архитектуры для звука • Трансформеры для звуковых данных. Модификации архитектуры «Трансформер» для обработки временных рядов звуков. Применение Audio Spectrogram Transformer. • **Лабораторная работа:** реализация модели «Трансформер» для задачи распознавания речи на наборе данных LibriSpeech. • Предобученные модели для обработки звука. Wav2Vec, HuBERT, Whisper: принципы работы, области применения. Тонкая настройка предобученных моделей для решения специализированных задач. • **Лабораторная работа:** дообучение Wav2Vec2 на задаче распознавания речи на пользовательском подмножестве данных.	**4**
Модуль 4. Прикладные задачи обработки звука • Распознавание речи (ASR). Полный пайплайн распознавания речи. Методы оценки: WER (Word Error Rate). • **Мастер-класс:** разработка системы ASR на основе Wav2Vec2 и/или Whisper. • Сегментация и выделение событий в аудио. Выделение звуковых событий из длинных записей. Применение моделей CRNN. • **Мастер-класс:** разработка модели для выделения событий на наборе данных UrbanSound8K. • Синтез речи (TTS). Архитектуры Tacotron, WaveNet. Совмещение синтеза речи с нейронными сетями. • **Мастер-класс:** разработка базовой TTS-системы на основе Tacotron2. • Устранение шумов и улучшение качества звука: Denoising Autoencoders, Conv-TasNet. Применение для очистки записей. • **Мастер-класс:** разработка модели удаления шума в наборе данных DNS Challenge.	**4**

Аудиторная нагрузка в классе с преподавателем	**16 ак.ч. + 16 ак.ч. самостоятельно**

По окончании обучения на курсе проводится итоговая аттестация. Аттестация проводится в виде теста на последнем занятии или на основании оценок практических работ, выполняемых во время обучения на курсе.

Контакты центра

Для частных лиц
+7 (495) 232-32-16

Для организаций
+7 (495) 780-48-44

E-mail
info@specialist.ru