

Распознавание человека по голосу

Магомедов Абдуррахман

Научный руководитель: (Ёжик | Александр Панин)

Цель

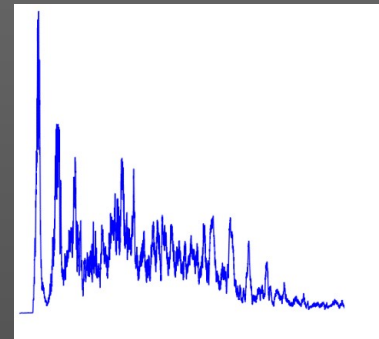
- Создать программу, распознающую человека по голосу и отличающую его от других людей

Задачи

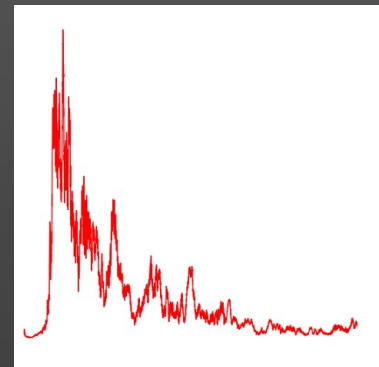
- Найти данные для обучения программы распознаванию человеческого голоса
- Выбрать подходящую модель для обучения
- Обучить модель
- Оценить качество работы модели

1 этап – обучение программы восприятию звука

- Для обучения программы использовался набор данных с сайта Open Speech and Language Resources (<http://www.openslr.org/12>)
- На первом этапе была поставлена промежуточная задача определения пола человека по голосу
- Программа обучалась на 5559 аудиозаписях английской речи



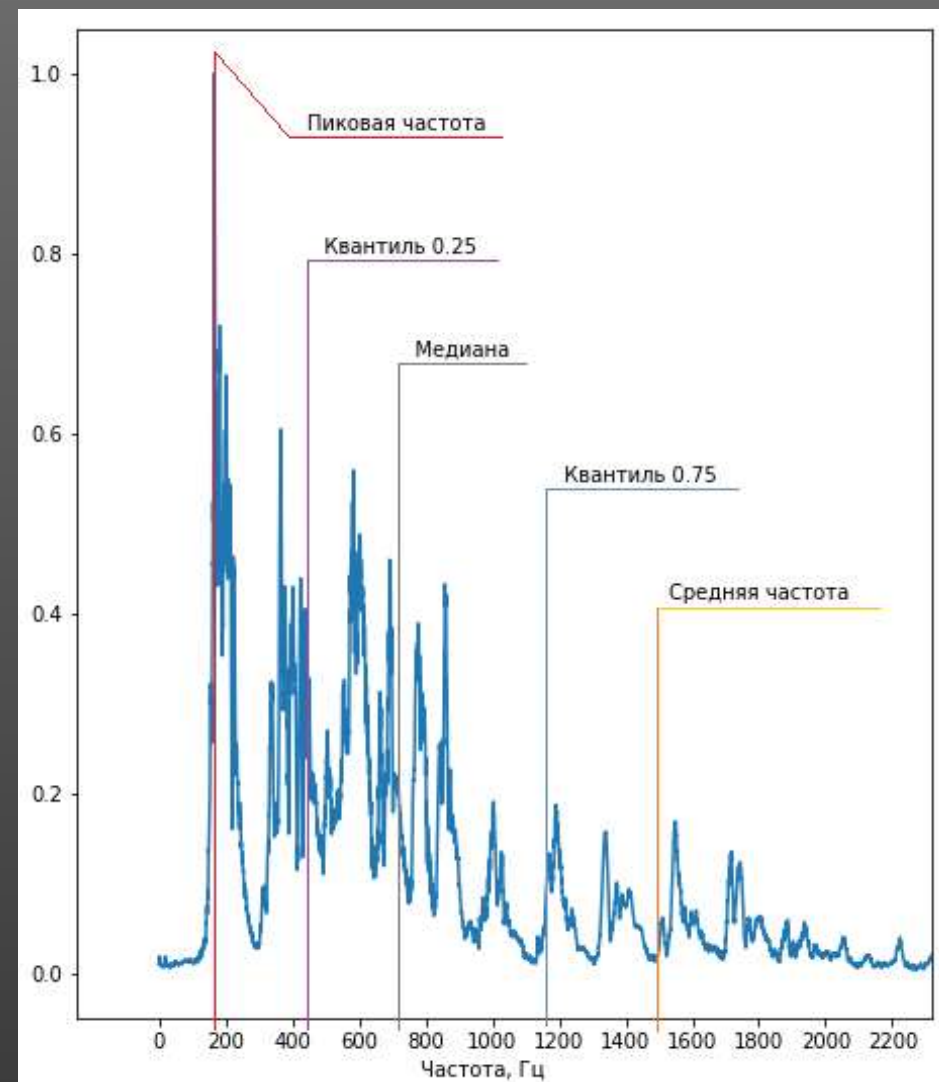
Мужской голос



Женский голос

Обработка звука

- Для решения задачи определения пола по голосу программа обрабатывает разложение звука на частоты
- Рассчитываются некоторые характеристики звука, например, средняя частота, медиана частот, дисперсия и т. д.



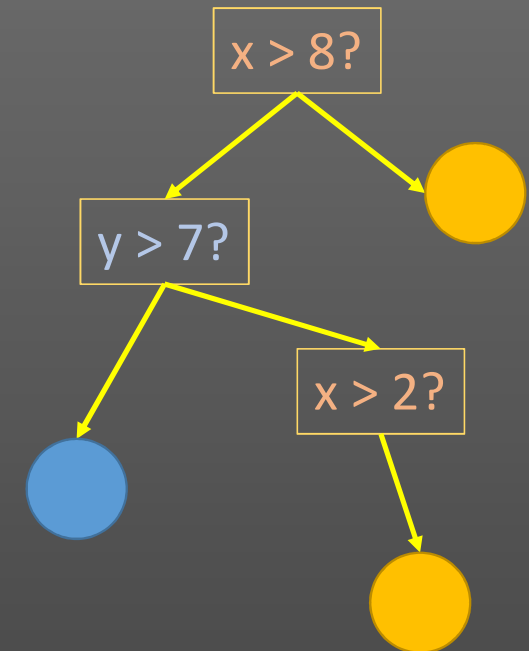
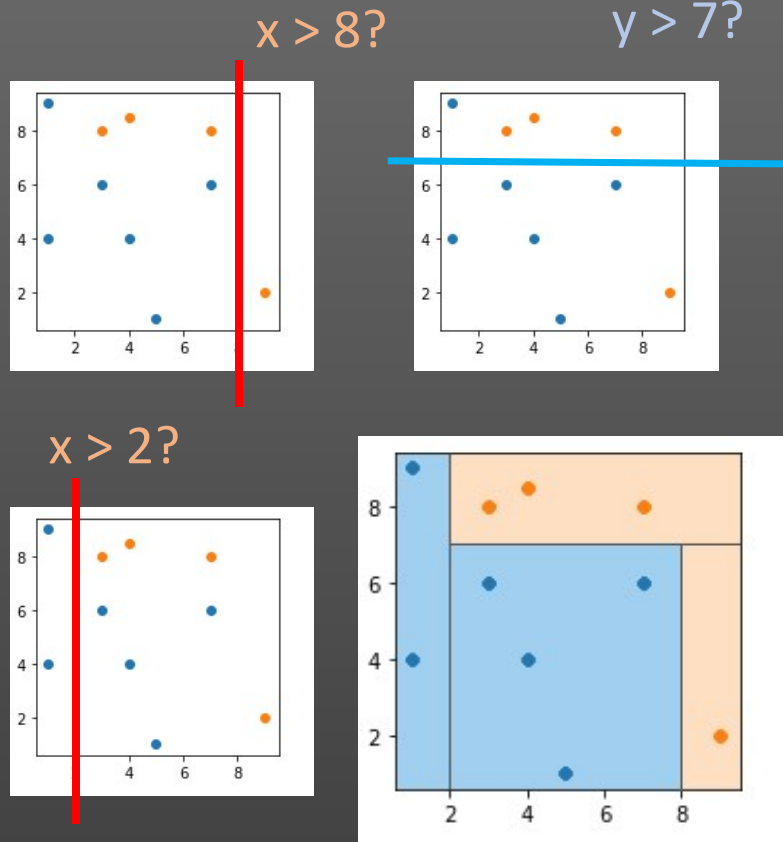
Обучающаяся модель – деревья решений

- В качестве базовой модели использованы деревья принятия решений
- Каждый узел дерева представляет вопрос о каком-либо признаке
- Для обучения необходим набор данных с заранее известным правильным ответом (характеристики звука и пол говорящего)



Обучение деревьев

- На каждом шаге выбирается какой-то признак, и по нему производится разделение таким образом, чтобы лучше всего разделить два класса



Достоинства

- Программа определяет пол говорящего в большинстве случаев

Недостатки

- Программа чувствительна к постороннему шуму
- Программа чувствительна к частому изменению интонации
- Время работы может сильно увеличиться в зависимости от размера файла

Что сделано на данный момент

- Набор данных (аудиозаписи с голосом людей) обработан, и получен массив характеристик для каждого аудиофайла
- Обработанные данные (характеристики звука + пол) разделены на две части – тестовую выборку (10%) и выборку для обучения
- Обученная модель показывает точность 99% на обучающей выборке, 86% на тестовой выборке, в среднем – 97%

Что нужно сделать

- Увеличить скорость работы программы
- Уменьшить влияние шума и изменения интонации
- Решить первоначальную цель – полноценное распознавание людей по голосу