

Санкт-Петербургский политехнический университет Петра Великого
Институт машиностроения, материалов и транспорта
Высшая школа автоматизации и робототехники

Курсовая работа

«Классификация голоса по полу записывающегося»
по дисциплине
«Математические методы интеллектуальных технологий»

Выполнил
студент гр. 3341506/00401

К.В.Потапов

Руководитель

А.В. Бахшиев

«___» _____ 2020 г.

Санкт-Петербург
2020

Содержание

ВВЕДЕНИЕ	3
1. Цели и задачи	4
1.1 Цели	4
1.2 Задачи	4
2. Теория.....	4
3. Анализ и обработка базы.....	5
3.1 Выбор базы	5
3.2 Анализ базы	6
3.3 Определения набора признаков для обучения.....	7
3.4 Постановка задачи машинного обучения.....	9
4. Применение алгоритмов машинного обучения	10
4.1 Выбор применяемых алгоритмов.....	10
4.2 Применение методов машинного обучения.....	10
ЗАКЛЮЧЕНИЕ	11
СПИСОК ЛИТЕРАТУРЫ.....	12

ВВЕДЕНИЕ

Машинное обучение использует сложные алгоритмы, чтобы «учиться» на массивах данных. Чем больше данных доступно алгоритмам, тем лучше они обучаются. Примеров работы машинного обучения в реальной жизни сегодня множество. Вспомните о персональных рекомендациях продуктов, распознавании лиц или мгновенно предлагаемых маршрутах картах.

Исторически сложилось так, что наибольшего успеха глубокое обучение достигло в задачах *image processing* – распознавания, сегментации и обработки изображений. Однако в этой работе попытаемся разобраться в задачах связанных с обработкой речи.

Самой популярной и востребованной из них является, вероятно, распознавание того, что именно говорят, анализ на семантическом уровне, но мы обратимся к более простой задаче – определению пола говорящего. Впрочем, инструментарий в обоих случаях оказывается практически одинаков.

1. Цели и задачи

1.1 Цели

В рамках настоящей курсовой работы рассматриваются следующие цели:

- изучение алгоритмов машинного обучения,
- изучение методов анализа баз данных для применения на них алгоритмов машинного обучения,
- применение вышеупомянутых алгоритмов на произвольном наборе данных.

1.2 Задачи

При выполнении работы затрагиваются следующие задачи:

- выбор произвольной базы данных,
- анализ выбранной базы и подготовка ее для применения алгоритмов машинного обучения,
- постановка задачи машинного обучения,
- применение алгоритмов машинного обучения
- анализ полученных результатов.

2. Теория

Первым делом нужно разобраться в физике процессов – понять, чем мужской голос отличается от женского. Голосовые связки, колебания которых производят звуковые волны до модуляции другими органами речи, у мужчин и женщин имеют различную толщину и натяжение, что ведет к разной частоте основного тона (она же *pitch*, высота звука). У мужчин она обычно находится в пределах 65-260 Гц, а у женщин – 100-525 Гц. Иными словами, мужской голос чаще всего звучит ниже женского. Видно, что для обоих полов эти интервалы довольно сильно перекрываются. Более того, в процессе речи

частоту основного тона – вариативный параметр, меняющийся, например, при передаче интонации, – невозможно определить для многих согласных звуков, а алгоритмы ее вычисления не безупречны. В восприятии человека индивидуальность голоса содержится не только в частоте, но и в тембре – совокупности всех частот в голосе.

3. Анализ и обработка базы

3.1 Выбор базы

В качестве базы данных взята база с сайта <https://data.world/ml-research/gender-recognition-by-voice>, основанная на 3168 звуковых дорожках, на которых записаны небольшие фразы людей. Эта база данных была создана для идентификации голоса как мужского или женского на основе акустических свойств голоса и речи. Набор данных состоит из 3168 записанных образцов голоса, собранных у говорящих мужчин и женщин. Голосовые образцы предварительно обрабатываются акустическим анализом в R с использованием пакетов seewave и tuneR с анализируемым частотным диапазоном от 0 до 280 Гц (диапазон человеческого голоса). Признаками являются: meanfreq - средняя частота (в кГц), sd - стандартное отклонение частоты, median - медианная частота (в кГц), Q25 - первый квантиль (в кГц), Q75 - третий квантиль (в кГц), IQR - межквантильный диапазон (в кГц), skew – асимметрия, Kurt - эксцесс, sp.ent - спектральная энтропия, SFM - спектральная плоскостность, mode - режим частоты, centroid - частота центроида, meanfun - средняя основная частота измеряется по акустическому сигналу, minfun - минимальная основная частота, измеренная по акустическому сигналу, maxfun - максимальная основная частота, измеренная по акустическому сигналу, meandom - средняя доминирующая частота, измеренная по акустическому сигналу, mindom - минимум доминирующей частоты, измеренная по акустическому сигналу, maxdom - максимум доминирующей частоты, измеренная по акустическому сигналу, dfrange - range доминирующей частоты, измеренной в акустическом

сигнале, modindx - индекс модуляции . Изначально по этой базе планируется предсказывать пол записывающегося.

3.2 Анализ базы

В таблице 1 представлены все признаки датасета .

Таблица 1 – Все признаки из датасета

Признак	Описание признака
meanfreq	оценивает среднюю частоту в терминах частоты (в кГц)
sd	Стандартное отклонение частоты
median	Медианная частота в (в кГц)
Q25	первый квантиль (в кГц)
Q75	третий квантиль (в кГц)
IQR	межквантильный диапазон (в кГц)
Skew	ассиметрия
Kurt	эксцесс
sp.ent	спектральная энтропия
SFM	спектральная плоскостность.
mode	режим частоты
centroid	частота центроида
meanfun	средняя основная частота измеряется по акустическому сигналу
minfun	минимальная основная частота, измеренная по акустическому сигналу
maxfun	максимальная основная частота, измеренная по акустическому сигналу
meandom	средняя доминирующая частота, измеренная по акустическому сигналу
mindom	минимум доминирующей частоты, измеренная по акустическому сигналу
maxdom	максимум доминирующей частоты, измеренная по акустическому сигналу
dfrange	range доминирующей частоты, измеренной в акустическом сигнале
modindx	индекс модуляции

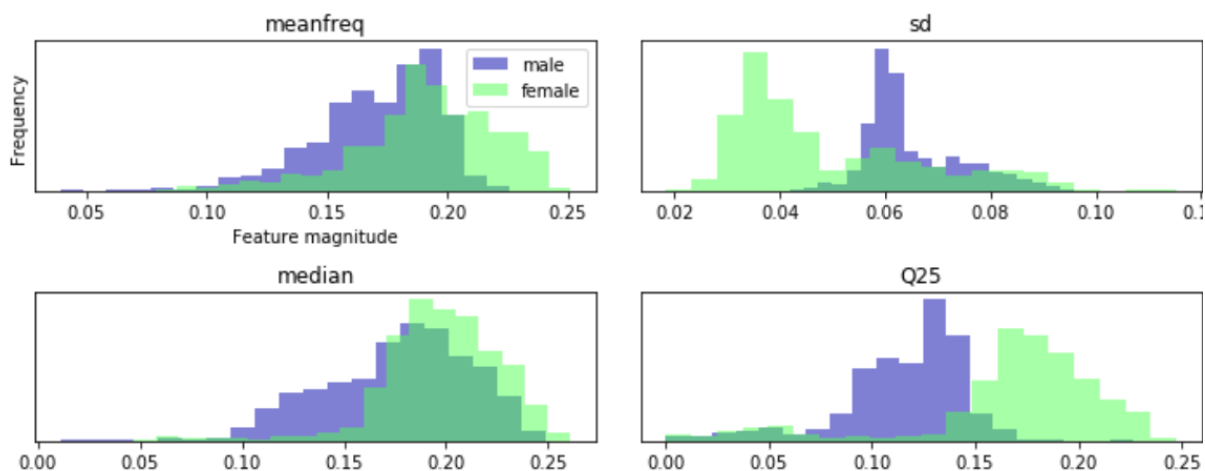
Таблица 2 – Анализ данных с помощью Pandas

	meanfreq	sd	median	Q25	Q75	IQR	skew	kurt	sp.ent	sfm	...
0	0.059781	0.064241	0.032027	0.015071	0.090193	0.075122	12.863462	274.402906	0.893369	0.491918	...
1	0.066009	0.067310	0.040229	0.019414	0.092666	0.073252	22.423285	634.613855	0.892193	0.513724	...
2	0.077316	0.083829	0.036718	0.008701	0.131908	0.123207	30.757155	1024.927705	0.846389	0.478905	...
3	0.151228	0.072111	0.158011	0.096582	0.207955	0.111374	1.232831	4.177296	0.963322	0.727232	...
4	0.135120	0.079146	0.124656	0.078720	0.206045	0.127325	1.101174	4.333713	0.971955	0.783568	...

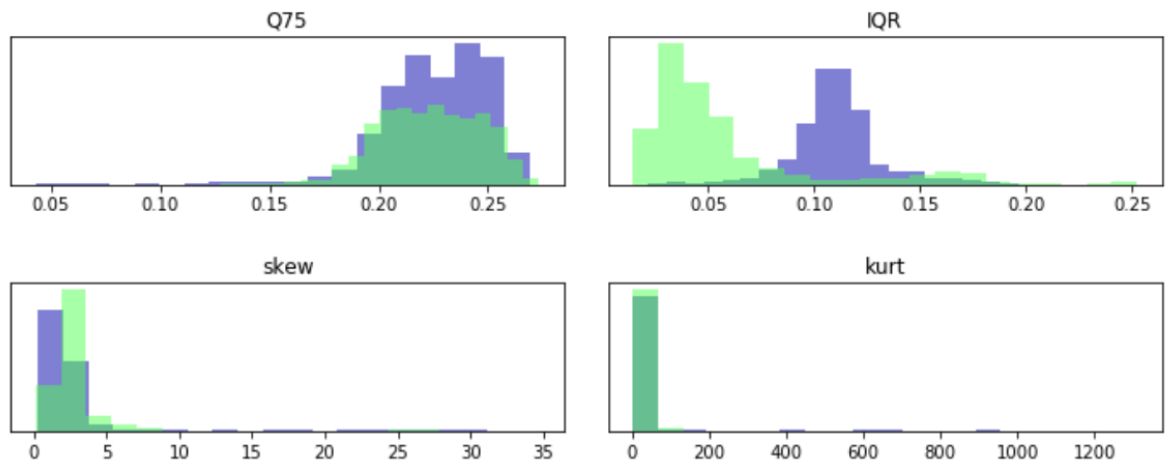
5 rows × 21 columns

3.3 Определения набора признаков для обучения

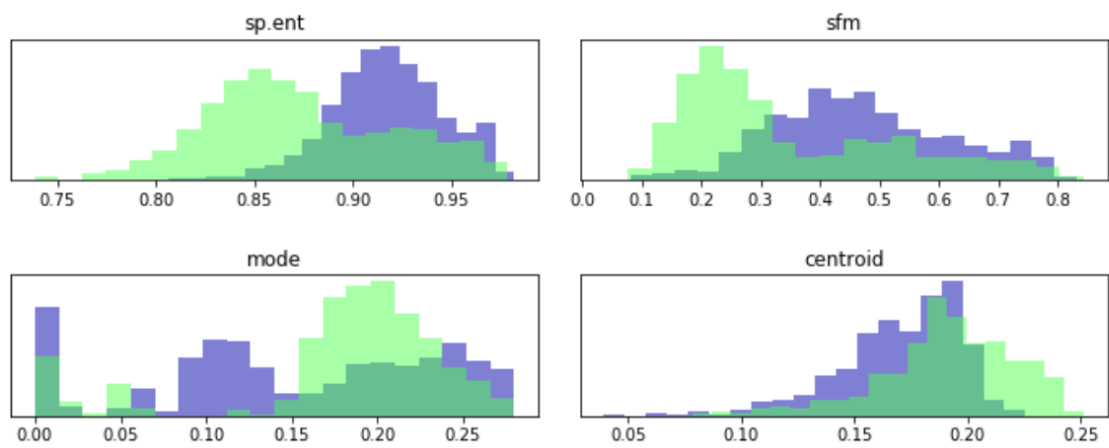
Далее для определения признаков, от которых сильно зависят целевые признаки были рассмотрены гистограммы между целевой и независимых переменных. Как мы видим ниже, такие переменные, как sd, Q25, IQR, sp.ent, sfm, mode и meanfun, могут помочь нам отделить мужские голоса от женских.



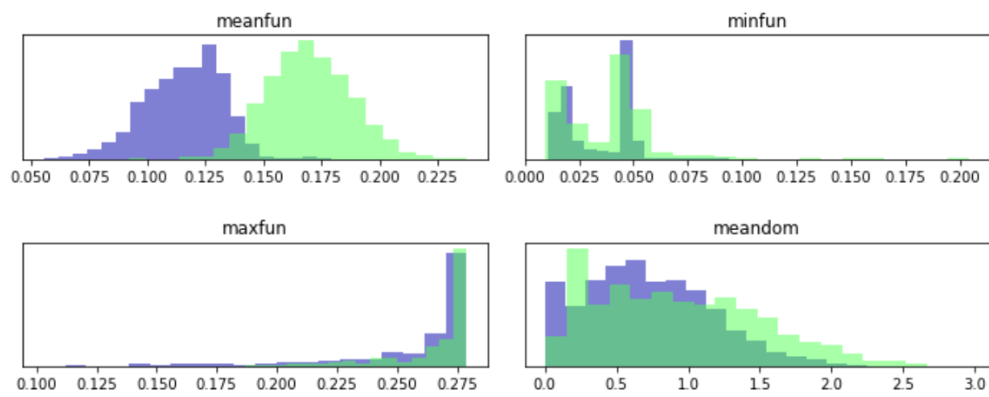
a) Meanfreq, sd, median, Q25



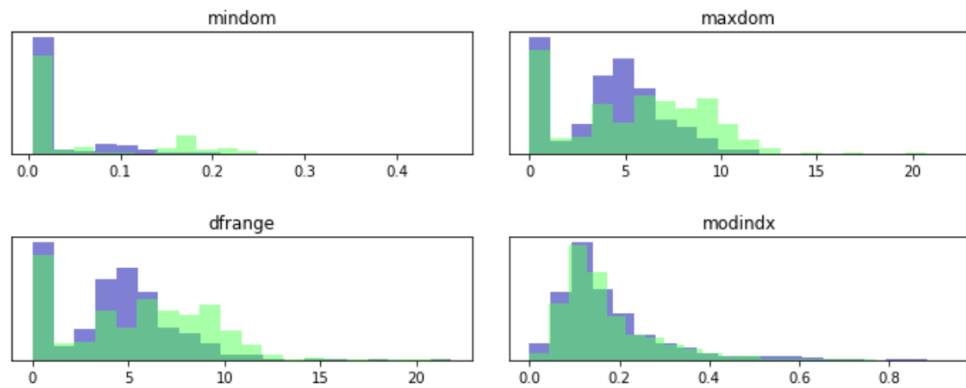
6) Q75, IQR, skew, kurt



B) sp.ent, mode, centroid, sfm



г) meanfun, minfun, maxfun, meandom



г) mindom, maxdom, dfrange, modindx

Рисунок 1 – Гистограммы целевой и независимой переменной

3.4 Постановка задачи машинного обучения

Для выбранной базы данных задача машинного обучения следующая: предсказание методом классификации пола записывающегося.

Решить задачу классификации – значит на обучающих данных построить функцию, которая по аналогичным параметрам возвращает метку класса, причем делает это довольно-таки точно. В нашем случае необходимо, чтобы по набору признаков для произвольного аудиофайла наш классификатор отвечал, чья речь в нем записана, мужчины или женщины.

4. Применение алгоритмов машинного обучения

4.1 Выбор применяемых алгоритмов

На основе прочитанной статьи [1], в которой лучше всего справился с поставленной задачей - случайный лес (Random forest). Поэтому для наших задач был выбран тоже случайный лес (Random forest). Кроме того, мы также разделим полный набор данных на наборы данных для обучения и тестирования.

4.2 Применение методов машинного обучения

В результате работы функции подбора параметров были получены файлы, в которых были записаны результаты предсказаний для метода классификации и типов входных данных. Данные из файлов для наилучшей читаемости были преобразованы в гистограмму.

Точность на обучающем наборе: 0,998

Точность на испытательном наборе: 0,976

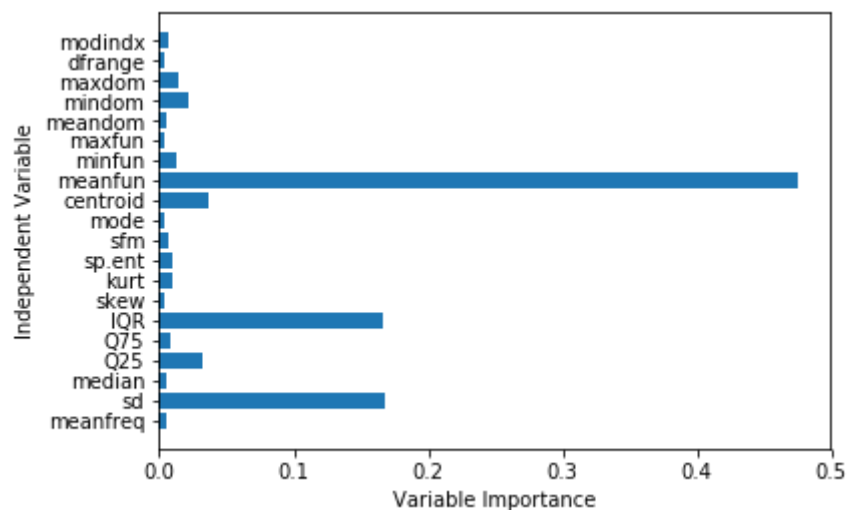


Рисунок 2 – Гистограмма полученных данных

ЗАКЛЮЧЕНИЕ

В рамках работы были изучены методы анализа и предварительной обработки баз данных, предоставляемых для алгоритмов машинного обучения. Найден произвольный датасет, для которого в дальнейшем была поставлена задача машинного обучения. Изучены и применены различные методы классификации. Машинное обучение в обработке речи – объективно сложно.

В результате проделанной работы была решена задача классификации определения пола по голосу записывающегося. Для алгоритма случайного леса значение процента верного предсказания достигло 97.6%. В целом данный алгоритм полностью удовлетворил поставленную задачу.

СПИСОК ЛИТЕРАТУРЫ

1. James Chen. Predict gender with voice and speech data – Текст: электронный // Интернет-портал. – URL: https://medium.com/@jameschen_78678/predict-gender-with-voice-and-speech-data-347f437fc4da
2. Стерлинг Г.,Казимирова Е. Случайный лес vs нейросети: кто лучше справится с задачей распознавания пола в речи data – Текст: электронный // Интернет-портал. – URL: <https://habr.com/ru/company/neurodatalab/blog/334136/>