

Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский университет «Высшая школа экономики»
Образовательная программа «Прикладная математика»
бакалавр

ОТЧЕТ

по проектной работе
Идентификация пользователя по голосу

Выполнили студенты группы БПМ 181
Бизьянов И.Ю.
Старшинов Ф.А.

Научный руководитель:
Иванов Фёдор Ильич

Москва
25 мая 2019 г.

Аннотация.

В данной работе представлен процесс построения модели для классификации пользователей на основе записей их голосов.

Код доступен по ссылке:

<https://github.com/IgorBizyanov/nis>

Обученная модель:

<https://yadi.sk/d/gxX1YQIp6tNmDQ>

Датасет:

<https://www.dropbox.com/sh/mfeocpk45c13rsb/AAAwYcJm639oe429tzwAKVHaa?dl=0>

Измененный модуль `audioSegmentation` библиотеки `pyAudioAnalysis`:

<https://github.com/IgorBizyanov/nis/blob/master/audioSegmentation.py>

Введение

Работа заключается в построении и реализации алгоритма распознавания пользователя по фрагменту его голоса с использованием известных методов машинного обучения и извлечения признаков из аудиосигнала.

Описание задачи и решение представлено в трёх главах:

- В первой главе опишем инструменты, используемые в ходе выполнения проектной работы;
- Во второй главе опишем процесс извлечения признаков, предобработки данных и построения модели;
- В третьей мы покажем, как меняется качество классификации в зависимости от параметров модели.

Заключение содержит выводы о проделанной работе.

ГЛАВА 1

В настоящий момент существует большое количество алгоритмов и методов машинного обучения. Каждый из них обладает известным набором достоинств и недостатков. Для выполнения этой работы был выбран оптимальный, на наш взгляд, метод *k*-ближайших соседей (*k*-nearest neighbors) с количеством "соседей" равным 13.

Для реализации классификатора был выбран язык Python, обладающий всем, что необходимо для продуктивной работы программиста и аналитика. Из всего множества библиотек для анализа данных и машинного обучения используются:

1. Numpy – поддержка высокоуровневых математических функций, предназначенных для работы с многомерными массивами;
2. sklearn – библиотека машинного обучения с множеством эффективно реализованных алгоритмов построения моделей;
3. pyAudioAnalysis (с небольшой модификацией одного из модулей) - библиотека, реализующая множество методов извлечения признаков из аудиозаписи, а также имеющая модули обработки данных;
4. Matplotlib – библиотека для построения графиков.

В качестве среды разработки программного обеспечения была выбрана оптимизированная для задач машинного обучения среда «Jupyter Notebook».

ГЛАВА 2

Любое обучение с учителем, как известно, нуждается в начальных размеченных данных, используя которые модель должна определять класс поданных ей на вход новых, неизвестных для нее ранее объектов. Датасет для обучения, ссылка на который указана в аннотации, включает более 2400 аудиозаписей, содержащих голоса разных дикторов, говорящих с изменчивой интонацией.

Для обучения модели необходимо иметь какие-либо числовые признаки, которые однозначно определяют поданные нашей модели на вход данные. Алгоритм извлечения этих признаков реализован в библиотеке `pyAudioAnalysis` в модуле `audioFeatureExtraction`. Максимальное количество таких признаков с точки зрения `pyAudioAnalysis` равно 34. Наша модель использует лишь первые 21 из них, так как это количество признаков является оптимальным по соотношению (качество модели)/(расход оперативной памяти)

Изначально, поданный библиотеке аудиофайл разделяет аудиозапись на небольшие фрагменты, именуемые фреймами, с определенной шириной окна и конкретным шагом, необходимым для указания степени перекрытия и наложения соседних фреймов друг на друга. Обучение модели происходит именно по фреймам, каждый из которых имеет метку своего класса.

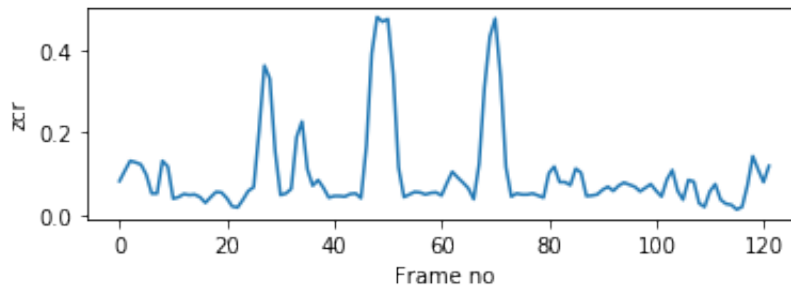


Рис. 1: Пример: извлечение признака zero-crossing rate.

Конечно, сырые данные зачастую содержат множество фоновых шумов, искажений и прочих помех, препятствующих повышению качества классификации. В нашем случае основными помехами являются паузы между произнесением слов (далее - тишина). Обучать модель на данных, имеющих тишину, по понятным причинам нерационально. К счастью, библиотека `pyAudioAnalysis` в модуле `audioSegmentation` содержит функцию `silenceRemoval`, возвращающая интервалы отсутствия тишины. По причине неудобства использования возвращаемых значений данная функция была несколько модифицирована, поэтому новая её версия возвращает индексы фреймов без тишины (для запуска модели необходимо лишь локально заменить файл модуля `audioSegmentation` на отредактированный; ссылка на измененный модуль - в аннотации). Удаление тишины происходит также с применением методов машинного обучения, однако в данном случае используется иная модель классификации - метод опорных векторов (`support vector machine`). Основным признаком здесь является энергия сигнала.

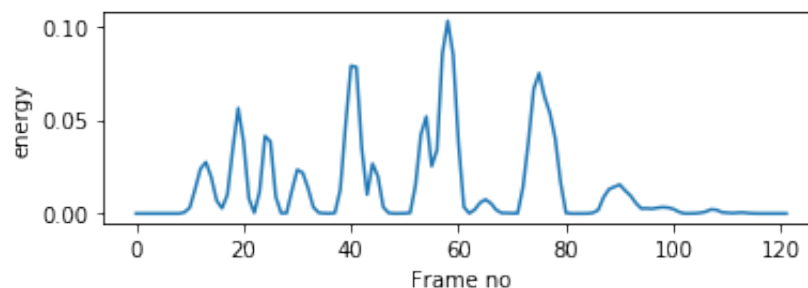


Рис. 2: Поведение признака energy.

ГЛАВА 3

Как уже было сказано, для обучения использован алгоритм билижайшего соседа. В ходе экспериментов, методом подбора, было выявлено количество "соседей при котором модель показывала наивысшее качество классификации объектов. В нашем случае наибольшая доля правильно классифицированных

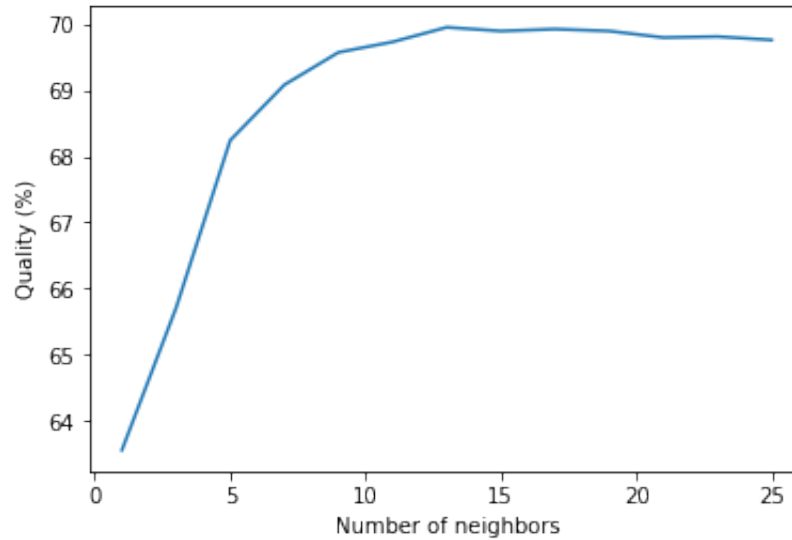


Рис. 3: Качество распознавания в зависимости от числа соседних векторов.

фреймов составляет 0.69950866, которой соответствует число соседей равное 13.

ЗАКЛЮЧЕНИЕ

Разумеется, качество классификации могло быть заметно улучшено с использованием дополнительных подходов к преобработке данных или вовсе иных методов машинного обучения. Однако результат, полученный нами в ходе исследования, является достаточным для того, чтобы с высокой вероятностью определять пользователя по аудиозаписи, имеющей низкое число помех.

СПИСОК ЛИТЕРАТУРЫ

1. <https://github.com/tyiannak/pyAudioAnalysis/wiki>
2. Ромацкий Д. Б. АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ МУЗЫКАЛЬНЫХ ПРОИЗВЕДЕНИЙ ПО ЖАНРАМ - 2017
3. Analysis and application of audio features extraction and classification method to be used for North Indian Classical Music's singer identification problem. - International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 2, February 2014
4. Understanding Machine Learning: From Theory to Algorithms - 2014 by Shai Shalev-Shwartz and Shai Ben-David