

Санкт-Петербургский государственный университет
Кафедра компьютерного моделирования и многопроцессорных систем

Мирошниченко Александр Сергеевич

Выпускная квалификационная работа бакалавра

Разработка системы распознавания речевых команд при помощи методов
машинного обучения

Направление 01.03.02

«Прикладная математика и информатика»

Научный руководитель,
кандидат физ.-мат. наук,
доцент
Козынченко В. А.

Санкт-Петербург

2021 г.

Содержание

Введение	3
Постановка задачи	4
Обзор литературы	5
Глава 1. Теоретические сведения	6
Глава 2. Описание решения	7
Глава 3. Результаты вычислений	8
Выводы	9
Заключение	10
Список литературы	11
Приложение	12

Введение

Постановка задачи

Обзор литературы

Глава 1. Теоретические сведения

Глава 2. Описание решения

Глава 3. Результаты вычислений

Было проведено 3 вычислительных эксперимента для каждого из 2-х типов нейронной сети: MLP и CNN. Структуры приведены на рис 1,2

Датасет состоит из 6 дикторов. Каждый диктор работал с 11 командами : 'back', 'down', 'menu', 'off', 'on', 'open', 'play', 'power', 'stop', 'up', 'volume'.

Диктор	Тип голоса	Кол-во звук. дорожек на каждую команду	Сумм. кол-во звук. дорожек
speaker1	Мужской	50	550
speaker2	Мужской	40	440
speaker3	Мужской	40	440
speaker4	Мужской	40	440
speaker5	Мужской	50	550
speaker6	Женский	50	550

Первый эксперимент: нейронная сеть обучается на первом дикторе с мужским голосом, тестирование производится на каждом дикторе.

Второй эксперимент: нейронная сеть обучается на всех дикторах с мужским голосом, тестирование производится на каждом дикторе.

Третий эксперимент: нейронная сеть обучается на всех дикторах, тестирование производится на каждом дикторе.

Датасет предварительно разделяется на тренировочную и тестовую части. На тренировочную часть отводится 70% данных диктора, на тестовую часть - 30%. В процессе тренировки после каждой эпохи тренировочные данные перемешиваются. 15% тренировочных данных в каждой эпохе - валидационные. В качестве метрики для оценки эффективности была выбрана метрика точности (accuracy), а для валидации - функция потерь категориальной кросс-энтропии (val_loss). Алгоритм оптимизации - Adam. Максимальное количество эпох - 50. Если значение метрики val_loss не уменьшается в течение 20 эпох, то обучение останавливается.

Графики обучения для каждого из экспериментов приведены на рисунках.

В конце каждого эксперимента помимо тестирования производится построение матрицы ошибок (confusion matrix) для каждого диктора и для каждого из четырех пороговых значений: 0.5, 0.6, 0.7, 0.8.

```
all_speakers = [speaker1, speaker2, speaker3, speaker4, speaker5, speaker6]
```

```
all_male_speakers = [speaker1, speaker2, speaker3, speaker4, speaker5]
```


train_data	test_speaker	cnn_loss	mlp_loss	cnn_accuracy	mlp_accuracy
speaker1	speaker1	0.057	0.171	0.982	0.97
speaker1	speaker2	12.073	7.279	0.182	0.152
speaker1	speaker3	4.392	2.325	0.348	0.409
speaker1	speaker4	10.642	8.054	0.235	0.197
speaker1	speaker5	2.963	4.176	0.606	0.539
speaker1	speaker6	17.923	7.528	0.139	0.073
all_male_speakers	speaker1	0.084	0.044	0.976	0.988
all_male_speakers	speaker2	0.051	0.184	0.985	0.939
all_male_speakers	speaker3	0.111	0.172	0.955	0.97
all_male_speakers	speaker4	0.714	0.771	0.932	0.894
all_male_speakers	speaker5	0.074	0.08	0.982	0.982
all_male_speakers	speaker6	9.217	9.326	0.321	0.267
all_speakers	speaker1	0.132	0.044	0.958	0.988
all_speakers	speaker2	0.084	0.224	0.977	0.909
all_speakers	speaker3	0.2	0.154	0.932	0.955
all_speakers	speaker4	0.863	0.57	0.894	0.902
all_speakers	speaker5	0.07	0.05	0.988	0.994
all_speakers	speaker6	1.734	2.267	0.709	0.618

Таблица 1: Результаты вычислений

Выводы

Заключение

В данной работе:

- Проведена предобработка звуковых дорожек, содержащих команды
- Разработан алгоритм распознавания речевых команд
- Реализован алгоритм распознавания речевых команд
- Проведены вычислительные эксперименты, в результате которых показана работоспособность и эффективность работы алгоритма распознавания речевых команд.

Список литературы

- [1] Aurélien G. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems / Aurélien G. — 2nd Edition — O'Reilly Media, 2019.
- [2] Kailash A. Generative Adversarial Networks Projects: Build next-generation generative models using TensorFlow and Keras / Kailash A. — Packt Publishing, 2019.
- [3] Документация TensorFlow [Электронный ресурс]. — Режим доступа: https://www.tensorflow.org/api_docs/python/tf
- [4] Портал ML Glossary [Электронный ресурс]. — Режим доступа: <https://ml-cheatsheet.readthedocs.io>
- [5] Курс на платформе Coursera [Электронный ресурс]. — Режим доступа: <https://www.coursera.org/learn/getting-started-with-tensor-flow2>

Приложение

Ссылка на репозиторий с кодом: <https://gitlab.com/polotent/boxy>