

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	3
Обзор литературы	3
Глава 1. Теоретические сведения	3
Глава 2. Описание решения	3
Глава 3. Результаты вычислений	3
Выводы	4
Заключение	5
Список литературы	6
Приложение	6

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 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.

Графики обучения для каждого из экспериментов приведены на рисунках.

Таблица с результатам

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>