Speech-Recognition

Exercise 02 Speech Recognition with Convolutional Neural Network, Part of LN / midterm, see iLearn course!

Prof. Dr.-Ing. Udo Garmann

Abstract

Learn how a Neural Network (NN) can be used for Speech Recognition

1. Prerequisites

Python knowledge

Jupyter Notebook (with venv)

You may do this exercise in a group of 2 students or alone.

2. Tutorial

With the following link, download and use the Jupyter Notebook. Then walk through the tutorial and answer the questions below.

Please change the seed to another value bigger than 42 and the number of epochs to bigger than 10 (but not too big, depends on your computer).

Create a text file with your answers and upload it in the iLearn course.

Tutorial:

https://www.tensorflow.org/tutorials/audio/simple_audio

3. Questions

Answer the following questions and create a text file for your answers as described in the next part:

1) What are the used (voice) commands?

2) Write down the value of your seed (change the default)!

3) What is the number of used samples?

4) Split filenames into training, validation and test sets. The default is a 80:10:10 ratio, respectively. Use a good variation of this default. What is your number of training files?

5) What is the name of the first file shown in the waveform diagram?

Hint to find the file: Look at the waveform diagrams of the commands. They appear in different order. What is the first command? In the data directory find the folder with the wave files of this command and get the first name of the file name list. Write this down with the word of first command, e.g. "The first command is 'go', the name of the first file is 0a9f9af7_nohash_0.wav".

6) Hear different WAV-files of a single command with an appropriate player-program. Describe the differences between the sounds you hear!

7) What is the sample rate for the WAV-file?

8) Experiment with value of EPOCHs and change the default value. Write down the value of your number of epochs.

9) What is the final test accuracy of your CNN?

10) Describe the classification process in your own words! What do you think are the features used for classification?

Description of the text file

Name your text file according to the following rule:

spr-ex2-lastname-firstname.md

with first- and lastname and mat-no of the groups' speaker in one line and first- and lastname and mat-no of the second student on the next line.

The content of the file must look like this:

```
lastname 1, firstname1, matno1

lastname 2, firstname2, matno2

1) What are the used (voice) commands?

(your list)

2) Write down the value of your seed:

(your value here)

3) What is the number of samples?

(your value here)

4) Split filenames into training, validation and test sets. The default is a 80:10:10
    ratio, respectively. Use a good variation of this default. What is your number of
    training files?

(your value here)

5) What is the name of the first file shown in the waveform diagram?

(your text here)

6) Hear different WAV-files with an appropriate player-program. Describe the differences
    between the sounds you hear and name at least two types of errors that may disturb
    the classification!

(your description here)

7) What is the sample rate for the WAV-file?

(your value here)

8) Experiment with value of EPOCHs and change the default value. Write down the value of
    your number of epochs.

(your value here)

9) What is the final test accuracy of your CNN?

(your value here)

10) Describe the classification process in your own words! What do you think are the
    features used for classification?
```

(your text here)

(your text here)