Data Science Lab: Process and methods
Politecnico di Torino

Project Assignment
Winter Call, A.Y. 2022/2023

Last update: January 5, 2023

1 Project dates

Start date: January 5, 2022 at 23:59 (CET)
Due date: February 1, 2022 at 23:59 (CET)

Due date is a **strict deadline**.

2 Problem description

Intent detection is a crucial task in the field of natural language processing (NLP) and speech recognition. It involves identifying the underlying intention or purpose of a spoken or written statement. This task is particularly important in the development of voice assistants and other audio-based applications, as it allows the system to understand and respond appropriately to the user's requests and commands.

In the context of audio intent detection, the system must be able to accurately interpret the user's spoken words and determine their intended meaning. This requires the system to have a deep understanding of language and context, as well as the ability to recognize and differentiate between different intents or purposes.

The importance of accurate audio intent detection cannot be overstated. It plays a vital role in the usability and effectiveness of voice assistants and other audio-based applications, and can greatly impact the user experience. A system that is unable to correctly interpret the user's intentions will be frustrating and difficult to use, leading to a poor user experience and potentially even causing the user to abandon the application altogether. On the other hand, a system that is able to accurately detect and respond to the user's intentions will be much more user-friendly and effective, leading to a better overall experience for the user.

For this project, you will work on an intent detection problem: given an input audio sample, your goal is to predict both the *action* requested and the *object* that is affected by the action.

2.1 Dataset

Warning: For this project, you are not allowed to use external datasets other than the one provided. Adoption of external resources will result in failure of the exam.

The dataset consists in a collection of audio file in a WAV format. Each record is characterized by several attributes. The following is a short description for each of them.

- *path*: the path of the audio file.

- *speakerId*: the id of the speaker.
- *action*: the type of action required through the intent.
- *object*: the device involved by intent.
- *Self-reported fluency level*: the speaking fluency of the speaker.
- *First Language spoken*: the first language spoken by the speaker.
- *Current language used for work/school*: the main language spoken by the speaker during daily activities.
- *gender*: the gender of the speaker.
- *ageRange*: the age range of the speaker.

An intent is given by the combination of an action with an object, therefore the information in the two respective columns must be combined to obtain the label to be used to address this task. The way this information should be combined is a simple string concatenation (e.g., if the action is "increase" and the object is "volume", the corresponding intent will be "increasevolume").

The dataset is located at:
https://drive.google.com/file/d/1gUPpqPTlgfzIyDU4eG6t5HoTlK5inLO4/view?usp=sharing

Within the archive, you will find the following elements:

- **audio**: a folder containing all the audio files in WAV format.
- **development.csv** (development set): a comma-separated values file containing the records from the development set. This portion does have the `action` and `object` columns, which you should use to obtain the labels to train and validate your models.
- **evaluation.csv** (evaluation set): a comma-separated values file containing the records corresponding to the evaluation set. This portion does not have the `action` and `object` columns.
- **sample_submission.csv**: a sample submission file.

2.2 Task

You are required to build a classification pipeline to predict the intent expressed in each audio recording in the Evaluation set.

2.3 Evaluation metric

Your submissions will be evaluated through Accuracy.

3 Submit your result

Submission file To get your results evaluated, you have to upload a result file on our submission competition platform, Data Science Lab Environment (DSLE). The submission file must be a `CSV` file formatted as follows:

```
Id,Predicted
0,increasevolume
1,decreasevolume
2,activatemusic
3,decreasevolume
...
```

The submission file must contains a header line and a row for each record in the Evaluation collection. Each row must have two fields:

- the `Id` field, which represent the path of the corresponding audio sample. It correspond to the column `path` in the evaluation CSV file.
- the `Predicted` label for the corresponding record.

You can find a sample submission file in the project material (see 2.1).

Submission platform The submission platform is the same one you used during the course laboratories. Therefore, you have to use your personal key. If you do not have the key or have problems using it, please send an email to `lorenzo.vaiani@polito.it`. Please refer to the guide on the course website, to go through the submission procedure.

You can find the DSLE platform at http://trinidad.polito.it:8888

4 Upload the report and the software

The report and the software have to be submitted by the due date reported in Section 1. This is a strict deadline.

Submission All the required files (i.e. for the report and the software) must be included in a **single ZIP archive**. The archive must be uploaded to the "Portale della Didattica", under the *Homework* section. Please use as description: **report_exam_winter_2023**.

Info: A ZIP archive is a ZIP archive, not a RAR, a 7z or, a tarball archive, nor any of those renamed with a trailing `.zip` extension.

Formatting rules The formatting rules for both the report and the software are described in the document with exam rules. You can find it on the course website.