MLPC Report - Task 3: Classification Experiments

Team OBSERVE

Johannes Grafinger **Jonas Gantar** **Leonhard Markus Spanring** **Reinhard Josef Pötscher**

Contributions

Johannes, Jonas and Reinhard were responsible for tasks 1.) Labeling Function, 2.) Data Split, 3.) Audio Features and 4.) Evaluation. Leonhard was responsible for task 5.) Experiments. All of us together were responsible for task 6.) Analysing Predictions. We prepared this report in the same constellation. We all worked together on the presentation. We held regular meetings at which each member presented their results and the others critically evaluated the work. Everyone communicated via a dedicated Discord server and the editing was done over a Github repository.

1.) Labeling Function

For your analysis, you may focus on a subset of the 58 classes provided.

a.) Assess how accurately the applied labeling functions capture the intended classes.

a.).1 Do the mapped classes correspond well to the free-text annotations?

a.).2 Are the labeled events clearly audible within the indicated time regions?

b.) Which audio features appear most useful for distinguishing between the classes of interest?

c.) How well do the chosen audio features group according to the discretized class labels?

c.).1 Do samples of the same class form tight clusters?

2.) Data Split

a.) Describe how you split the data for model selection and performance evaluation.

b.) Are there any potential factors that could cause information leakage across the data splits if they are not carefully designed?

b.).1 If yes, how did you address these risks?

c.) Describe how you obtained unbiased final performance estimates for your models.

3.) Audio Features

a.) Which subset of audio features did you select for your final classifier?

a.).1 Describe the selection process and the criteria you used to make your choice.

b.) Did you apply any preprocessing to the audio features?

b.).1 If so, explain which techniques you used and why they were necessary.

4.) Evaluation

a.) Which evaluation criterion did you choose to compare hyperparameter settings and algorithms, and why?

b.) What is the baseline performance? What could be the best possible performance?

5.) Experiments

a.) For at least three different classifiers, systematically vary the most important hyperparameters and answer the following questions for each of them:

a.).1 How does classification performance change with varying hyperparameter values? Visualize the change in performance.

a.).2 (To what extent) Does overfitting or underfitting occur, and what does it depend on?

b.) After selecting appropriate hyperparameters, compare the final performance estimate of the three classifiers.

6.) Analysing Predictions

Find two interesting audio files that have not been used for training and qualitatively evaluate your classifier's predictions.

a.) Use the spectrogram and the sequence of predictions to visualize the classifier output.

b.) Listen to the audios and inspect the corresponding predictions of the classifier.

b.).1 How well does the classifier recognize the classes?

c.) What are particular problematic conditions that cause the classifier to mispredict classes?

c.).1 Can you think of simple postprocessing steps that might help improve the predictions?