MLPC Report - Task 2: Data Exploration

Team OBSERVE

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

Contributions

Reinhard and Johannes (Group 1) were responsible for tasks 1.) Case Study and 2.) Annotation Quality
Leonhard and Jonas (Group 2) were responsible for tasks 3.) Audio Features and 4.) Text Features of the report
All of us together were responsible for task 5.) Conclusions
In the same constellation we created this report. We all worked on the presentation together, with Group 2 providing its content. We held regular meetings, where each group presented their results up to that point and the other critically reviewing their work.

1.) Case Study

To find 2 interesting records that were edited by multiple annotators, we first looked in "metadata.csv" to see which files had more than one annotator. This resulted in a list of 149 files. We then looked at the "metadata$_{t}itle_{e}mbeddings.npz''andthe``metadata_{k}eywords_{e}mbeddings.npz''inordertobeabletodrawsomeconclusions.Atthesa$
$1.0)inoneplace)leadtoveryclearannotations.Animportantassumptionhereisthecorrectnessandaccuracyofthetitlesandkeywo$

a.) Identify similarities or differences between temporal and textual annotations from different annotators.

b.) To what extent do the annotations rely on or deviate from keywords and textual descriptions in the audio's metadata?

c.) Was the temporal and text annotations done according to the task description?

2.) Annotation Quality

a.) How precise are the temporal annotations?

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

b.) How similar are the text annotations that correspond to the same region?

c.) How many annotations did we collect per file? How many distinct sound events per file?

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

d.) How detailed are the text annotations? How much does the quality of annotations vary between different annotators?

e.) Are there any obvious inconsistencies, outliers, or poor-quality annotations in the data? Propose a simple method to filter or fix incorrect or poor-quality annotations (e.g., remove outliers, typos, or spelling errors).

3.) Audio Features

a.) Which audio features appear useful? Select only the most relevant ones or perform a down projection for the next steps.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

b.) Extract a fixed-length feature vector for each annotated region as well as for all the silent parts in between. The most straightforward way to do this is to average the audio features of the corresponding region over time, as shown in the tutorial session.

c.) Cluster the audio features for the extracted regions. Can you identify meaningful clusters of audio features? Do the feature vectors of the silent regions predominantly fall into one large cluster?

4.) Text Features

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

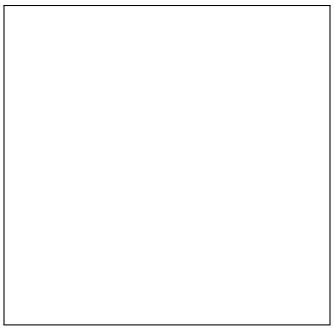

Figure 1: Sample figure caption.

a.) Cluster the text features. Can you find meaningful clusters?

b.) Design a labeling function1 for classes dog and cat. Do the annotations labeled as dog or cat sounds form tight clusters in the text and audio feature space?

c.) How well do the audio feature clusters align with text clusters?

5.) Conclusions

a.) Is the dataset useful to train general-purpose sound event detectors?

b.) Which biases did we introduce in the data collection and annotation phase?

6.) Submission of MLPC reports

Please read the instructions below carefully and follow them faithfully. Note that this template is based on the official Neurips 2023 template. In your report, you may use three levels of headings, as described in what follows.

7.) Headings: first level

This is a first level heading.

a.) Headings: second level

This is a second level heading.

a.).1 Headings: third level

And this is a third level heading. Make sure to structure your report s.t. no deeper levels are necessary.

8.) Footnotes, Figures and Tables

a.) Footnotes

Footnotes should be used sparingly. Note that footnotes are properly typeset *after* punctuation marks.[1]

b.) Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. You may use color figures. Please refer to all your figures in text, by using e.g., Figure 1.

[1] As in this example.

Table 1: Sample table title

	Part	
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

c.) Tables

All tables must be centered, neat, clean and legible. Please refer to all your tables in text, by using e.g., Table 1.

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package.[2]

9.) Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (this will result in a deduction of points). Please note that pages should be numbered, and adhere to the given *page limit* to avoid further point deductions. Your final submission should be a `pdf` file.

[2] `https://www.ctan.org/pkg/booktabs`