Worksheet 6

Conor Houghton, copying from Martha Lewis

1 Introduction

This worksheet is a little different from the previous ones: there is no programming! Instead, the idea is for you to read about a dataset and create a datasheet for it, using the ideas from Gebru et al. [2].

We will be using the datasheet questions from Gebru et al. [2], and applying them to the ImageNet dataset [1] to create a datasheet for the ImageNet dataset. If you cannot find the answers to all the questions, don't worry! The idea is for you to start thinking about the ideas and also to see the extent to which the information needed for the datasheet is publicly available.

2 Task

First of all, open up the paper *Datasheets for Datasets*. The URL is here:

`https://arxiv.org/abs/1803.09010`.

The set of questions that the authors propose should be used is in Sections 3.1-3.7. In each section, pick two or three of the questions and try to answer them in relation to the ImageNet dataset (paper here:

`http://www.image-net.org/papers/imagenet_cvpr09.pdf`

and website here: `http://www.image-net.org/`).

At the end of the *Datasheets for Datasets* paper, there are two example datasheets that you can look at to get ideas. Feel free to work together if it is helpful!

- How easy is it to answer the questions?

- Is one category of questions easier to answer than another?

- How easy would it be to find out whether the dataset is suitable for a project you want to carry out?

- Do you identify any aspects of the dataset that you think should be altered?

- Any other points you want to make?

References

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. *ImageNet: A large-scale hierarchical image database.* In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 248–255, 2009.

[2] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. *Datasheets for datasets.* Communications of the ACM, 64(12):86–92, 2021. URL: `https://arxiv.org/abs/1803.09010`.