Evaluating data reduction techniques for supervised training

Yuwen Heng

Master of Science
Data Science, Technology, and Innovation
School of Informatics
University of Edinburgh
2020

Abstract

Training deep neural networks can be resources-consuming. The budget required are increasing with the size of the dataset. During the past few decades, many research is dedicated to develop training procedures to accelerate the convergence speed of deep learning. However, we still need the whole dataset to train the network and paying for a large dataset may not pay back well if we can use a smaller subset to achieve an acceptable performance. To solve this issue, we first adapted and evaluated three methods, Patterns by Ordered Projections (POP), Enhanced Global Density-based Instance Selection (EGDIS), and Curriculum Learning (CL), to reduce the size of two image datasets, CIFAR10 and CIFAR100, for the classification task. Based on the analysis, we present our two contributions: the Weighted Curriculum Learning (WCL) and a trade-off framework. The WCL outperforms POP and EGDIS in terms of both classification accuracy and time complexity. It achieves comparable performance compared with CL while keeping a portion of hard examples. The trade-off framework selects a subset of samples according to the acceptable relative accuracy and the dataset. In addition, the framework is also extended to predict the amount of samples needed to achieve a particular accuracy with a given subset.

Acknowledgements

Any acknowledgements go here.

Any acknowledgements go here.

Table of Contents

Chapter 1

Introduction

Besides, paying for a large dataset may not pay back well if we can use a much smaller one to achieve an acceptable performance. The preliminary material of your report should contain:

- The title page.

- An abstract page.

- Optionally an acknowledgements page.

- The table of contents.

    As in this example `skeleton.tex`, the above material should be included between:

```
\begin{preliminary}
    ...
\end{preliminary}
```

This style file uses roman numeral page numbers for the preliminary material.

The main content of the dissertation, starting with the first chapter, starts with page 1. ***The main content must not go beyond page 40.***

The report then contains a bibliography and any appendices, which may go beyond page 40. The appendices are only for any supporting material that's important to go on record. However, you cannot assume markers of dissertations will read them.

You may not change the dissertation format (e.g., reduce the font size, change the margins, or reduce the line spacing from the default 1.5 spacing). Over length or incorrectly-formatted dissertations will not be accepted and you would have to modify your dissertation and resubmit. You cannot assume we will check your submission

before the final deadline and if it requires resubmission after the deadline to conform to the page and style requirements you will be subject to the usual late penalties based on your final submission time.

1.1 Using Sections

Divide your chapters into sub-parts as appropriate.

1.2 Citations

Citations (such as [1] or [2]) can be generated using `BibTeX`. For more advanced usage, the `natbib` package is recommended. You could also consider the newer `biblatex` system.

These examples use a numerical citation style. You may also use (Author, Date) format if you prefer.

Chapter 2

Your next chapter

A dissertation usually contains several chapters.

Chapter 3

Conclusions

Chapter 4

New file

I want to know what is it.

4.1   Final Reminder

The body of your dissertation, before the references and any appendices, *must* finish by page 40. The introduction, after preliminary material, should have started on page 1.

You may not change the dissertation format (e.g., reduce the font size, change the margins, or reduce the line spacing from the default 1.5 spacing). Over length or incorrectly-formatted dissertations will not be accepted and you would have to modify your dissertation and resubmit. You cannot assume we will check your submission before the final deadline and if it requires resubmission after the deadline to conform to the page and style requirements you will be subject to the usual late penalties based on your final submission time.

Bibliography

[1] Hiroki Arimura. Learning acyclic first-order Horn sentences from entailment. In *Proc. of the 8th Intl. Conf. on Algorithmic Learning Theory, ALT '97*, pages 432–445, 1997.

[2] Chen-Chung Chang and H. Jerome Keisler. *Model Theory*. North-Holland, third edition, 1990.