Qeios

Peer Review

Review of: "Optimizing Multi-GPU Training with Data Parallelism and Batch Size Selection"

Zhenguo Ma[1]

1. Zhejiang Lab, Hangzhou, China

Through empirical analysis, the paper investigates the effect of different batch sizes on training efficiency and model performance, providing valuable insights into how to select the optimal batch size for large-scale datasets in multi-GPU environments. However, several aspects of the paper require further refinement.

1. Novelty of the Paper

The novelty of this paper appears to be somewhat limited. It is important to note that a substantial body of research already addresses batch size optimization in multi-GPU environments, such as studies [1], [2], and [3]. These works have thoroughly examined the impact of batch size on training efficiency and model performance in similar contexts.

2. Organization of the Paper

The overall organization of the paper requires substantial improvement in the following areas:

- Both the Abstract and the Introduction fail to clearly articulate how this work distinguishes itself from existing studies.
- The discussion in Section 3.2 regarding the "Justification for MobileNetV2" seems unnecessary and does not significantly contribute to the core argument of the paper.
- A considerable portion of Section 4.2 is devoted to discussing existing work, which would be more appropriately placed in Section 2 (Related Work).
- The conclusion in Section 5 is somewhat verbose and could be made more concise.

3. Experimental Section

Several concerns regarding the experimental section require clarification and improvement:

- The authors determine the 'optimal' batch size by empirically comparing different batch sizes without presenting a theoretical analysis. This empirical approach may be less convincing without a solid theoretical basis.
- The accuracy difference between the single-GPU (0.26) and multi-GPU (>0.9) setups is notably large. I suggest the authors provide a more detailed explanation for this substantial discrepancy.
- The validation loss curve for the multi-GPU setup shows an upward trend, which is atypical and potentially problematic. Further investigation is needed to understand this result.
- It would be helpful to present the training time for different batch sizes to allow for a comparative analysis of training efficiency.

doi.org/10.32388/P1G574

- On page 8, the results for batch size 64 are presented in Figure 4, rather than Figure 5 as indicated in the paper. The reference should be corrected for consistency.

- The results for batch size 128 appear to be missing from the paper. If these results were part of the study, they should be presented and discussed.

4. Minor Issues

There are a few additional minor issues that should be addressed to enhance the overall quality and professionalism of the paper:

- The capitalization of section titles is inconsistent. For instance, "3.3.1. Single-GPU Training" and "3.3.2. Multi-GPU Training" follow different capitalization styles. A consistent formatting style should be applied throughout the manuscript.

[1] AdaBatch: Adaptive Batch Sizes for Training Deep Neural Networks. arXiv preprint arXiv:1712.02029, 2017.

[2] Improving Scalability of Parallel CNN Training by Adjusting Mini-Batch Size at Run-Time. IEEE International Conference on Big Data (Big Data), 2019.

[3] Dynamic Interplay of Mini-Batch Size and Aggregation Frequency for Federated Learning with Static and Streaming Dataset. IEEE Transactions on Mobile Computing, 2023.

Declarations

Potential competing interests: No potential competing interests to declare.