Hybrid Optimization Strategy for Enhanced Deep Learning on MNIST Dataset

Sourav Panda, Srivatsa Bhamidipati, Shubhang Kaushik

Abstract

This work introduces a hybrid optimization approach that combines Adam's efficiency with SGD's robust generalization properties. Customized transition conditions are tailored for specific tasks, datasets, and models. Experiments on MNIST and diverse models show improved accuracy, convergence speed, and generalization. This approach bridges the gap between fast progress and strong generalization in deep learning, with implications for future advancements.

1 Introduction

Deep learning has revolutionized artificial intelligence, enabling machines to excel in complex tasks with unmatched precision and efficiency. The optimization algorithm governing deep neural network training is pivotal for their success. Stochastic Gradient Descent (SGD) [1] is a cornerstone in this regard, but it has its limitations, especially regarding learning rate determination, making training challenging and time-consuming.

The choice of optimization algorithm profoundly impacts both model performance and training efficiency in deep learning. Adaptive techniques like Adaptive Moment Estimation (Adam) [2] have gained prominence, with extensive applications across various machine-learning domains [3]. Nonetheless, recent research has raised concerns about adaptive methods matching SGD in terms of generalization performance [4].

2 Related Work

The field of deep learning optimization has witnessed substantial research and innovation, driven by the need to improve training efficiency and model performance. In this section, we discuss key developments in optimization techniques for deep neural networks.

2.1 Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent (SGD) is a fundamental optimization algorithm in deep learning, known for its simplicity and effectiveness. However, it is highly sensitive to hyperparameters, especially the learning rate. Choosing the right learning rate can be challenging, as an incorrect choice may lead to slow convergence or divergence during training. Researchers have worked on mitigating this challenge through learning rate schedules and adaptive methods.

2.2 Adaptive Optimization Methods and Their Limitations

In response to SGD's limitations, adaptive optimization methods have gained prominence. They dynamically adjust learning rates, aiming to speed up convergence and eliminate the need for manual learning rate adjustments. One notable method in this category is Adam. However, recent research has raised concerns about adaptive methods' generalization performance compared to SGD [4]. Studies suggest that adaptive methods may struggle to match SGD's generalization abilities, especially in complex tasks. This has led to growing interest in hybrid optimization strategies that combine adaptive techniques' initial benefits with SGD's superior generalization properties.

2.3 Transitioning from Adaptive to SGD

The concept of shifting from adaptive optimization methods to SGD during training is not new and has been explored in different contexts. For example, [5] employed a mixed strategy, fine-tuning the transition point and learning rate for SGD post-transition. Likewise, [6] used a hybrid approach, combining RMSProp and SGD steps with a convex combination. These studies have paved the way for investigating hybrid optimization strategies that smoothly transition between adaptive methods and SGD.

3 Proposed Solution

3.1 Methodology

Our project aims to adopt an approach that combines the advantages of Adam's swift initial progress with the robust generalization properties of Stochastic Gradient Descent (SGD), as introduced in the research paper [7]. This transition occurs seamlessly, guided by predefined conditions tailored to the specific task at hand.

3.2 Tailored Conditions for Transition

A crucial aspect of our solution is the precise definition of conditions that trigger the shift from the adaptive method to SGD. These conditions will be customized to the specific task, dataset, and model architecture under consideration, ensuring that the transition occurs at the most opportune moments to improve generalization performance.

3.3 Experimental Validation on MNIST and Model Architectures

To empirically validate the effectiveness of our hybrid optimization strategy, we will rigorously apply it to the MNIST dataset, a renowned benchmark for image classification tasks. We will conduct a series of experiments, meticulously measuring performance metrics such as accuracy, convergence speed, and generalization performance. Furthermore, our research extends beyond MNIST. We plan to apply this methodology to different model architectures, encompassing scenarios of under-fit, good-fit, and over-fit. This comprehensive investigation will provide valuable insights into the versatility and efficiency of the hybrid optimization approach across various learning regimes.

4 Project Timeline

- **Week 1-2:** Literature Review and Requirements Analysis
  - Conduct a literature review on optimization algorithms (Adam, SGD) and transition strategies.
  - Download the MNIST dataset and explore the data.
- **Week 3-4:** Hybrid Approach Development
  - Develop the hybrid optimization algorithm (Adam to SGD transition).
  - Integrate the hybrid approach into deep learning frameworks.
- **Week 5-6:** Testing and Evaluation on the MNIST dataset
  - Test the hybrid approach on the MNIST dataset.
  - Evaluate performance, convergence speed, and generalization against standalone Adam and SGD.
- **Week 7-8:** Testing and Evaluation of different model architectures
  - Test the hybrid approach on different model architectures (underfit, good fit, and overfit).
  - Evaluate the findings, deduce insights, and finalize the project report.

5 Expected Outcomes

Expected outcomes include improved accuracy and training efficiency on the MNIST dataset and different model architectures through our hybrid optimization strategy. We also anticipate gaining valuable insights into the strategy's effectiveness when compared to traditional methods, potentially influencing the future of deep learning optimization.

6 Conclusion

This project presents a new hybrid optimization approach that improves the performance of deep learning on the MNIST dataset and its efficiency in addressing issues with different model architectures. To find a good mix between fast progress and strong generalization, we want to use the best features of adaptive optimization methods like Adam along with the dependability of SGD. Our carefully thought-out project schedule makes sure that tasks are done in a logical order and that full results paperwork is made. This study adds important new information to the field of deep learning optimization by showing how well hybrid methods work.

References

[1] Robbins, Herbert, and Monro, Sutton. "A stochastic approximation method." The Annals of Mathematical Statistics 22.3 (1951): 400-407.

[2] Kingma, Diederik P., and Ba, Jimmy. "Adam: A method for stochastic optimization." International Conference on Learning Representations (ICLR), 2015.

[3] Karpathy, A. "A Peek at Trends in Machine Learning." https://medium.com/@karpathy/apeek-at-trends-in-machine-learningab8a1085a106, 2017. [Online; accessed 12-Dec2017].

[4] Wilson, A. C., Roelofs, R., Stern, M., Srebro, N., and Recht, B. "The Marginal Value of Adaptive Gradient Methods in Machine Learning." ArXiv e-prints, May 2017.

[5] Wu, Y., Schuster, M., Chen, Z., Le, Q., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. "Google's neural machine translation system: Bridging the gap between human and machine translation." arXiv preprint arXiv:1609.08144, 2016.

[6] Akiba, T., Suzuki, S., and Fukuda, K. "Extremely large minibatch SGD: Training resnet-50 on ImageNet in 15 minutes." arXiv preprint arXiv:1711.04325, 2017.

[7] Nitish, K., Richard, S., "Improving Generalization Performance by Switching from Adam to SGD." arXiv preprint arXiv:1712.07628, 2017.