MLA Collection: Nesterov-accelerated Adaptive Moment Estimations

NAdam - V01 (30/05/2024)

Juliao Braga[1], Jeferson Campos Nobre[2], Itana Stiubiener[1], Juliana Cristina Braga[1], Percival Henriques[3], and Gemini Advanced[4]

[1]UFABC
[2]UFRGS
[3]ANID
[4]Google

1 Introduction

Nesterov-accelerated Adaptive Moment Estimation (NAdam) is an optimization algorithm used in machine learning to update network weights during training. It combines the benefits of two other optimization methods, Adam and Nesterov accelerated gradient (NAG).

2 Purpose

The primary purpose of NAdam is to achieve faster convergence and improved performance compared to its predecessors. It does this by adapting the learning rate for each parameter (Adam) and using momentum to guide the direction of updates (NAG).

3 Approach

NAdam incorporates the momentum term into the gradient update rule, providing a "look ahead" mechanism to anticipate future gradients. This helps accelerate convergence, especially in situations with sparse gradients.

4 Benefits

- Faster convergence compared to standard stochastic gradient descent.
- Handles sparse gradients effectively.
- Adapts learning rates for individual parameters.

5 Disadvantages

- May require hyperparameter tuning for optimal performance.
- Can be computationally expensive compared to simpler methods.

6 Variations

There aren't many significant variations of NAdam, as it is already a combination of two existing methods. However, researchers continue to explore modifications to further improve its performance.

7 Applications

NAdam is widely used in various machine learning tasks, including training deep neural networks for image recognition, natural language processing, and other complex problems.

8 Learning Types and Machine Learning Methods

NAdam is primarily used for supervised learning tasks, where the goal is to learn a mapping from input data to output labels.

NAdam is commonly used within the gradient descent framework for training various machine learning models, including deep neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs).

Figure 1 displays a knowledge graph that could be matched with the ontology proposed in the OBMLA-IID project (Ontology-based Machine Learning Algorithms for the Internet Infrastructure Domain) [1].

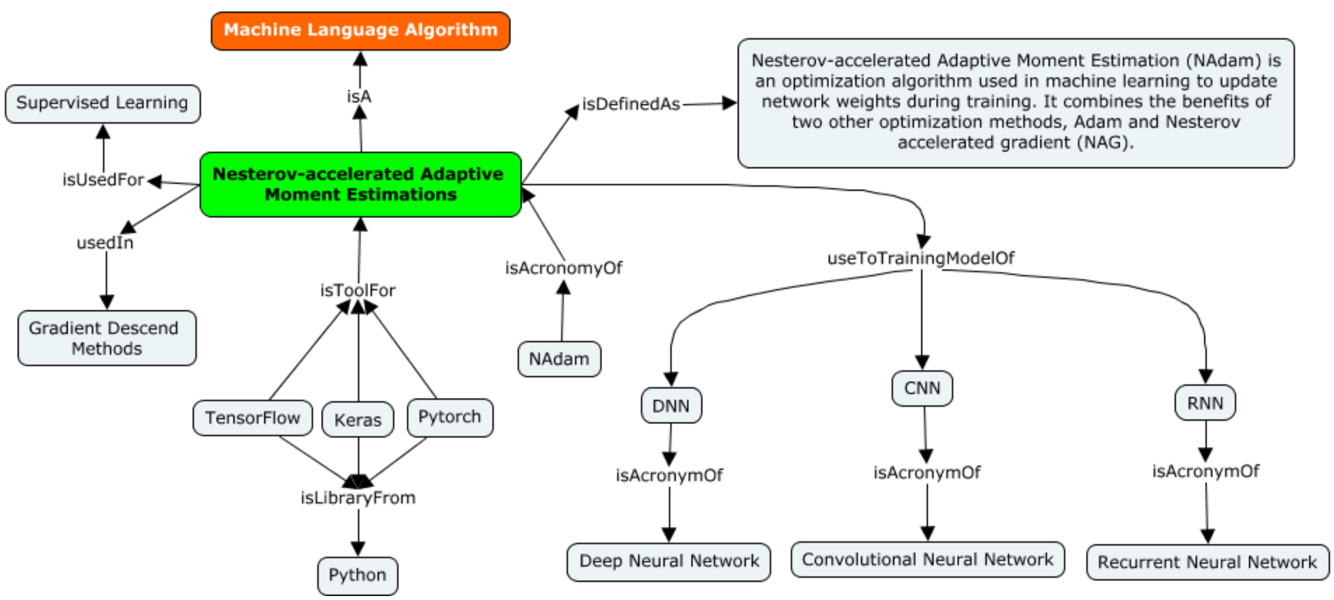

Figure 1: Knowledge graph of the main components of the algorithm, to match with an Ontology.

9 Libraries and Tools

- Python: TensorFlow, Keras, PyTorch
- Other Languages: Implementations are available in other languages like C++, Java, and R.

10 Challenges

The main challenge in using NAdam lies in selecting appropriate hyperparameters, such as learning rates and momentum coefficients. These may need to be tuned for specific datasets and tasks to achieve optimal performance.

11 Key Reference

[2]

12 Acronym: NAdam

13 Final considerations

This document is part of the ODBMLA-IID Project's collection of Machine Learning algorithms and briefly describes the Nesterov-Accelerated Adaptive Moment Estimations (NAdam) algorithm. Relevant part of this document was obtained using the following prompt related to Google's Gemini Advanced:

To the Machine Language algorithm: Nesterov-accelerated Adaptive Moment Estimation. Give me a summary description of no more than 10 lines. Characterise it including purpose, approach, benefits, disadvantages, variations and applications. Give the learning types and explain me how the algorithm is used on each. In the context of Machine Learning give me the methods that this algorithm use. Inform the libraries and tools available in each computer languages. Display challenges in its use, if applicable. Give a key reference. If exist, give me its acronym. Give me you answer in English, please.

References

[1] Juliao Braga, Itana Stiubiener, Jeferson C Nobre, Juliana C Braga, and Percival Henriques. OBMLA-IID: Ontology-based Machine Learning Algorithms for the Internet Infrastructure Domain, May 2024. DOI: `https://doi.org/10.17605/OSF.IO/GXRA7`.

[2] Timothy Dozat. Incorporating Nesterov Momentum into Adam. *ICLR Workshop*, 1:2013–2016, 2016.