

Accepted Manuscript

Distributed and Asynchronous Stochastic Gradient Descent with Variance Reduction

Yuewei Ming, Yawei Zhao, Chengkun Wu, Kuan Li, Jianping Yin

PII: S0925-2312(17)31803-9
DOI: [10.1016/j.neucom.2017.11.044](https://doi.org/10.1016/j.neucom.2017.11.044)
Reference: NEUCOM 19109

To appear in: *Neurocomputing*

Received date: 28 September 2016
Revised date: 9 October 2017
Accepted date: 19 November 2017

Please cite this article as: Yuewei Ming, Yawei Zhao, Chengkun Wu, Kuan Li, Jianping Yin, Distributed and Asynchronous Stochastic Gradient Descent with Variance Reduction, *Neurocomputing* (2017), doi: [10.1016/j.neucom.2017.11.044](https://doi.org/10.1016/j.neucom.2017.11.044)



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Distributed and Asynchronous Stochastic Gradient Descent with Variance Reduction

Yuewei Ming^{a,*}, Yawei Zhao^a, Chengkun Wu^a, Kuan Li^a, Jianping Yin^b

^aCollege of Computer,

National University of Defense Technology, Changsha 410073, China

^bState Key Laboratory of High Performance Computing,

National University of Defense Technology, Changsha, 410073, China

Abstract

Stochastic Gradient Descent (SGD) with variance reduction techniques has been proved powerful to train the parameters of various machine learning models. However, it cannot support the distributed systems trivially due to the intrinsic design. Although conventional studies such as PetuumSGD perform well for distributed machine learning tasks, they mainly focus on the optimization of the communication protocol, which does not exploit the potential benefits of a specific machine learning algorithm. In this paper, we analyze the asynchronous communication protocol in PetuumSGD, and propose a distributed version of variance reduced SGD named DisSVRG. DisSVRG adopts the variance reduction technique to update the parameters in a model. After that, those newly learned parameters across nodes are shared by using the asynchronous communication protocol. Besides, we accelerate DisSVRG by using the adaptive learning rate with an acceleration factor. Additionally, an adaptive sampling strategy is proposed in DisSVRG. The proposed methods greatly reduce the wait time during the iterations, and accelerate the convergence of DisSVRG significantly. Extensive empirical studies verify that DisSVRG converges faster than the state-of-the-art variants of SGD, and gains almost linear speedup in a cluster.

Keywords: Stochastic gradient descent, Variance reduction, Asynchronous

* Corresponding author

Email address: ywming@nudt.edu.cn (Yuewei Ming)

Download English Version:

<https://daneshyari.com/en/article/6864619>

Download Persian Version:

<https://daneshyari.com/article/6864619>

[Daneshyari.com](https://daneshyari.com)