PERUI: A General Framework of Reduced Variance Stochastic Gradient Gradient and the Hybrid Implementation

Yawei Zhao
School of Computer
National University
of Defense Technology
Changsha, China, 410073

Yuewei Ming
School of Computer
National University
of Defense Technology
Changsha, China, 410073

Jianping Yin
School of Computer
National University
of Defense Technology
Changsha, China, 410073

Algorithm 1 The general framework of variance reduced SGD: PERUI

Require: $\omega^0 \in \mathbb{R}^d$. $\forall i \in [n]$, and $[n]$ represents $\{1, 2, ..., n\}$.
Probability: $[i_t] \leftarrow \mathcal{P}([n])$ where $i_t \in \{1, 2, ..., n\}$. t is a positive integer;
Epoch: the sequence of the epoch size $\{m^0, m^1, ..., m^S\} \leftarrow \mathcal{E}([i_t])$;
for $s = 0, 1, 2, ..., S$ **do**
  $\omega_0^s = \tilde{\omega}^s$;
  $g = \frac{1}{n} \sum i = 1^n \nabla f_i(\tilde{\omega}^s)$;
  for $t < m^s$ **do**
    Reduced variance: $v = \mathcal{R}(\nabla f_{i_t}(\omega_{i_t}^t) - \nabla f_{i_t}(\tilde{\omega}^s))$;
    $\gamma_t^s = v + g$;
    Update: $\omega_{t+1}^s = \mathcal{U}(\eta_t, \omega_t^s, \gamma_t^s)$;
  Identification: $\tilde{w}^{s+1} \leftarrow \mathcal{I}([\omega_j^s])$ with $j \in \{1, 2, ..., m^s\}$;
  return $\tilde{\omega}^S$.

Abstract

Introduction

Related work

The general hybrid framework of SGD

Convergence analysis

Optimization

Constant learning rate with an acceleration factor

Adaptive update sharing strategy

Discussion

Performance evaluation

Convergence

Speed up

Wait time

Parallel threads

Conclusion

Acknowledgements