

Incremental Consensus based Collaborative Deep Learning

A short report on [?]

Nazarov Ivan

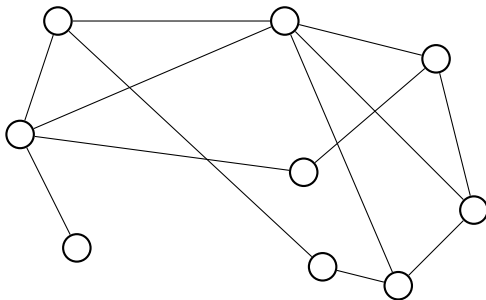
Skoltech

July 27, 2018

General Setting

Agents connected within a communication graph

- ▶ want to collaboratively solve an optimization problem



Generic Optimisation in ML

Agents are nodes in a undirected connected graph $G = (V, E)$

Each node $v \in G$ has a set of objectives $S_v = (l_{vi}: \mathbb{R}^d \rightarrow \mathbb{R})_{i=1}^{m_v}$

- ▶ $l_{vi}(\theta) = \text{Loss}(x_{vi}, y_{vi}; \theta)$ – supervised learning
- ▶ $l_{vi}(\theta) = \text{Loss}(z_{vi}; \theta)$ – unsupervised learning

Goal – to solve the global problem

$$\underset{\theta \in \Theta}{\text{minimize}} \quad \sum_{v \in G} \sum_{i \in v} l_{vi}(\theta). \quad (\text{Gbl-P})$$

Distributed Optimisation in ML

Assumptions on G

- ▶ two-way communication: if $uv \in E$ then $vu \in E$
- ▶ for any $v \in G$ the neighbours $G_v = \{u \in G : vu \in E\}$
- ▶ single connected component

Assumptions on l_{vi} : γ_{vi} -smooth, proper and coercive

- ▶ $l_{iv}(y) \leq l_{iv}(x) + \langle \nabla l_{iv}(x), y - x \rangle + \frac{\gamma_{vi}}{2} \|y - x\|^2$ for all y
- ▶ $\text{dom } l_{vi} = \{\theta : l_{vi}(\theta)\} \neq \emptyset$
- ▶ coercive $l_{vi}(\theta) \rightarrow \infty$ as $\|\theta\| \rightarrow \infty$

Distributed and Decentralized Optimisation

Consensus problem

$$\begin{aligned} & \underset{\theta_v \in \Theta}{\text{minimize}} && \sum_{v \in G} \sum_{i \in v} l_{vi}(\theta_v). \\ & \text{subject to} && \theta_v = \theta_u, \forall v \in G \ \forall u \in G_v. \end{aligned} \quad (\text{Cons-P})$$

Equivalent problem: if A – symmetric adjacency matrix of G then

$$\begin{aligned} & \underset{\theta \in \Theta^G}{\text{minimize}} && \sum_{v \in G} \sum_{i \in v} l_{vi}(\theta_v). \\ & \text{subject to} && A\theta = \theta. \end{aligned}$$

The Optimization Updates in Distributed Optimization

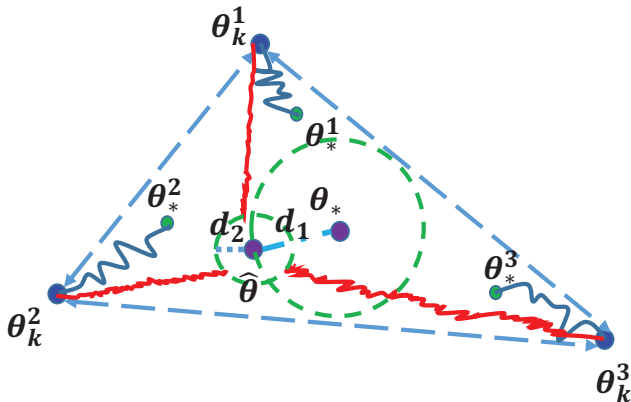


Figure 1: **Blue** dots ($\theta_k^i \sim \theta_v^t$) represent the current states; **Green** dots represent the private local optima ($\theta_*^i \sim \theta_v^*$); **Purple** dot ($\theta_* \sim \theta^*$) represents the ideal global optimal solution; another **purple** dot $\hat{\theta} \sim \hat{\theta}^t$ represents a possible consensus point.

Consensus Distributed SGD

CD-SGD update looks like this

$$\boldsymbol{\theta}_v^{t+1} \leftarrow \underbrace{\left\{ \frac{1}{|G_v|} \sum_{u \in G_v} \boldsymbol{\theta}_u^t \right\}}_{\text{synchronisation step}} - \underbrace{g_v^t(\boldsymbol{\theta}_v^t) \eta}_{\text{gradient step}},$$

where $g_v^t(\boldsymbol{\theta}_v^t)$ is the stochastic gradient of l_v at $\boldsymbol{\theta}_v^t$

$$g_v^t(\boldsymbol{\theta}_v^t) = \nabla_{\theta} \hat{\mathbb{E}}_{l \sim S_v} l(\theta) \Big|_{\theta = \boldsymbol{\theta}_v^t}.$$

Proposal of the paper

Instead of this ... (CDSGD)

$$\boldsymbol{\theta}^{t+1} \leftarrow (\boldsymbol{\Pi} \otimes I_d) \boldsymbol{\theta}^t - \mathbf{g}(\boldsymbol{\theta}^t) \eta.$$

... why not do this

$$\boldsymbol{\theta}^{t+1} \leftarrow (\boldsymbol{\Pi}^\tau \otimes I_d) \boldsymbol{\theta}^t - \mathbf{g}(\boldsymbol{\theta}^t) \eta,$$

thereby increasing the communication complexity by $\tau \geq 1$.

... and get

- ▶ the *incremental*-CDSGD (i-CDSGD)

Results

Provide vanilla and momentum variants of Gradient Descent iterations

Prove convergence for $\tau \geq 1$ using Lyapunov stability analysis

Results: consensus bound

Consensus with fixed step size, i-CDSGD

If assumptions hold, then iterates of i-CDSGD satisfy the following inequality $\forall t \geq 1$, when $0 < \alpha \leq \frac{1}{\gamma_m}$,

$$\mathbb{E} \left\| \theta_v^t - \frac{1}{|G|} \sum_{v \in G} \theta_v^t \right\| \leq \frac{\omega \alpha h}{1 - \hat{\lambda}_2}, \quad (1)$$

where $s^t = \frac{1}{N} \sum_{v \in G} \theta_v^t$, $\hat{\lambda}_2$ is the second-largest eigenvalue of the matrix $\mathbf{Q} = (1 - \omega)\mathbf{P} + \omega I_{Nd}$, $\gamma_m = \max_{v \in G, i \in v} \gamma_{vi}$, B_m - upper bounds the variance of the.

Slide with bullet points

This is a bullet list of two points:

- ▶ Point one
- ▶ Point two

Slide with two columns

Text goes in first column.

Text goes in second column

Slide with table

Table 1: Caption for table one

Heading1	Heading2	Heading3
Row1	0.1	.01
Row2	0.2	.02
Row3	0.3	0.03
Row4	0.4	0.04

Slide with figure



Figure 2: Caption for figure one.

Slide with references

This is to reference a figure (Figure 2)

This it to reference a table (Table 1)

This is to cite an article [Ahmed and Kim, 2018]

This is to add an article to the references without mentioning in the text

References I



Ahmed, M. and Kim, D. (2018).

pcr: An R package for quality assessment, analysis and testing of qPCR data.
PeerJ, 2018(3).