Robust Asynchronous and Network-Independent Cooperative Learning

Eduardo Mojica-Nava
Universidad Nacional de Colombia
eamojican@unal.edu.co

David Yanguas-Rojas
Universidad Nacional de Colombia
dryanguasr@unal.edu.co

César A. Uribe
LIDS-MIT
cauribe@mit.edu

Abstract

We consider the model of cooperative learning via distributed non-Bayesian learning, where a network of agents tries to jointly agree on a hypothesis that best described a sequence of locally available observations. Building upon recently proposed weak communication network models, we propose a robust cooperative learning rule that allows asynchronous communications, message delays, unpredictable message losses, and directed communication among nodes.

1 Robust Asynchronous Cooperative Learning: Algorithm and Main Result

Distributed inference has gained increasing attention in recent years due to the numerous applications in machine learning, sensor networks, decentralized control, and distributed signal processing. Among distributed inference models, non-Bayesian social learning has emerged as an essential approach to deal with decentralized heterogeneous learning over networks [3, 5]. Non-Bayesian learning exhibits strong theoretical performance and allows large classes of sensing modalities and communication constraints. The non-Bayesian learning model assumes that the network of agents tries to agree on a set of beliefs about the state of the world that best describes a sequence of local observations from a finite set of possible states [3, 5, 4, 7, 6]. In this work, we build upon recently available results in distributed optimization considering asynchrony, delays, and message losses [10, 9, 2, 1], and introduce a cooperative distributed non-Bayesian learning algorithm with robust performance guarantees under such harsh communication network conditions. In particular, we extend the recent proposed Robust Asynchronous Push-Sum (RAPS) consensus algorithm [8] to the distributed learning setup. Consider a network of n agents on a set of nodes $V = \{1, 2, \dots, n\}$ observing realizations of a finite, stationary, independent, identically distributed random processes $\{X_k\}_{k\geq 1}$ where $X_k^i \sim P^i$ at each iteration time k with unknown distribution P^i. Additionally, all agents have a shared finite set of hypotheses $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$, from which each agent $i \in V$ defines a local family of distributions $\mathcal{P}^i = \{P_\theta^i \mid \theta \in \Theta\}$. We will assume the technical condition that each element in family of distributions $\mathcal{P}^i$ is absolutely continuous with respect to P^i. We denote $\mathcal{N}_i^+$ and $\mathcal{N}_i^-$ as the set of out-neighbors and in-neighbors of an agent i. The objective of the network of agents is to agree on a parameter $\theta^* \in \Theta$ such that the joint distribution $\prod P_{\theta^*}^i$ is closest to $\prod P^i$. Formally, the group of agents tries to solve jointly: $\min_{\theta\in\Theta} F(\theta) \triangleq \sum_{i\in V} D_{KL}(P^i \| P_\theta^i)$, where $D_{KL}(P\|Q)$ is the Kullback-Leibler divergence between the distributions P and Q. Importantly, note that each of the agents only knows its local family of distributions $\mathcal{P}^i$, the true distribution of their local observations P^i is unknown, yet accessible via local observations. Thus, in order to solve problem cooperation is needed.

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

We will generally denote the set of minimizers as Θ^*. The confidence each agent has on each of the hypotheses in Θ is represented by a *belief* vector, denoted as $\mu_i^\theta(k)$, which indicates the belief that an agent $i \in V$ has about a hypothesis $\theta \in \Theta$ at certain time instant k. A value of $\mu_i^\theta(k) = 1$ indicates certainty that the minimizer is θ, whereas $\mu_i^\theta(k) = 0$ indicates certainty that it is not. Agents cooperate by communicating their beliefs at each time instant. Such communication is mediated by a network, modeled as a graph $\mathcal{G} = \{V, E\}$. We assume that the graph $\mathcal{G}$ is strongly connected and does not have self-loops, the delays on each link are bounded above by some $L_{del} \geq 1$, every agent wakes up and performs updates at least every $L_u \geq 1$ iterations, each link fails at most $L_f \geq 1$ consecutive times, and messages arrive in the order of time they were sent. We state Algorithm 1, our proposed cooperative learning algorithm, and state our main results.

Algorithm 1 Robust Asynchronous Push-Sum Distributed Non-Bayesian Learning

Initialize: $y_i(0) = 1$, $\phi_i^y(0) = 0$, $\phi_{i,\theta}^\mu(0) = 1$, $\forall i \in V$, and $\rho_{ij}^y(0) = 0$, $\kappa_{ij}(0) = 0, \forall (i,j) \in \mathcal{E}$
Set initial beliefs as uniform for all agents.
for $k = 0, 1, 2, \ldots,$ for every node i: **do**
  if Node i wakes up **then**
1. Processing and broadcasting local information
    $\kappa_i \leftarrow k$, $\phi_i^y \leftarrow \phi_i^y + y_i/(d_i^+ + 1)$
    $\phi_{(i,\theta)}^\mu \leftarrow \phi_{(i,\theta)}^\mu \left(\mu_i^\theta\right)^{y_i/(d_i^+ + 1)}$
    Node i broadcasts $(\phi_i^y, \phi_{(j,\theta)}^\mu, \kappa_i)$ to $\mathcal{N}_i^+$.
2. Processing received messages
    for $(\phi_j^y, \phi_{(i,\theta)}^\mu, \kappa_j')$ in the inbox **do**
      if $\kappa_j' > \kappa_{ij}$ **then**
        $\rho_{ij}^{*y} \leftarrow \phi_j^y$, $\rho_{ij|\theta}^{*\mu} \leftarrow \phi_{(j,\theta)}^\mu$, $\kappa_{ij} \leftarrow \kappa_j'$
      end if
    end for
3. Updating beliefs and local information
    $\hat{y}_i \leftarrow \frac{y_i}{d_i^+ + 1} + \sum_{j \in \mathcal{N}_i^-} (\rho_{ij}^{*y} - \rho_{ij}^y)$
    $\mu_i^\theta \leftarrow \frac{1}{Z_i} \left(\left(\mu_i^\theta\right)^{\frac{y_i}{d_i^+ + 1}} \prod_{j \in \mathcal{N}_i^-} \left(\frac{\rho_{ij|\theta}^{*\mu}}{\rho_{ij|\theta}^\mu} \right) P_\theta^i(x_{k+1}^i) \right)^{\frac{1}{\hat{y}_i}}$
Z_i is a normalization constant.
    $y_i \leftarrow \hat{y}_i$, $\rho_{ij}^y \leftarrow \rho_{ij}^{*y}$, $\rho_{ij|\theta}^\mu \leftarrow \rho_{ij|\theta}^{*\mu}$
  end if
end for

In Algorithm 1, each awake node executes three main states at every iteration. Initially, local variables are updated with the most recent information about outgoing neighbors for each possible hypothesis. This local processing step is concluded by broadcasting auxiliary variables and time-stamps to its available out-neighbors at that particular time. Then, each agent modes on processing the messages it might have arrived from its in-neighbors while not awake. Each agent first checks time-stamps for each of the messages and updates the stored neighbor information if newer information is available. Finally, the node updates its beliefs with the most recent information from its neighbors, and its local observation of the random variable X_k^i, and goes back into sleep mode. This process repeats at each iteration. We show that the learning dynamics proposed in Algorithm 1 guarantees that the beliefs of all agents will concentrate in the set of minimizers of $F(\theta)$, denoted as Θ^*. Additionally, the concentration rate is network-independent, in the sense that as the connectivity assumptions holds, the concentration rate will not depend on the specific network topology.

Theorem 1 (Main Result) *Let assumptions about the connectivity of the communication network hold. Then, the output of Algorithm 1 has the following property:*

$$\lim_{k \to \infty} \frac{1}{k} \log \frac{\mu_{\theta_v}^i(k)}{\mu_{\theta_w}^i(k)} \leq -\frac{1}{n} \min_{\theta \notin \Theta^*} (F(\theta) - F(\theta^*)) \tag{1}$$

almost surely for all $\theta_v \notin \Theta^*$, and $\theta_w \in \Theta^*$, and $i \in V$.

Theorem 1 states that for all non-optimal hypothesis, the beliefs will decay asymptotically exponentially fast. Moreover, the rate at which the beliefs will asymptotically decay is upper bounded by the averaged optimality gap of the second-best hypothesis. However, the *closer* (in the sense of Kullback-Leibler) the optimal and the closest suboptimal hypothesis are, the slower the concentration will happen.

References

[1] Christoforos N Hadjicostis, Alejandro D Domínguez-García, and Themistokis Charalambous. Distributed averaging and balancing in network systems. *Foundations and Trends in Systems and Control*, 5(2-3):99–292, 2018.

[2] Christoforos N Hadjicostis, Nitin H Vaidya, and Alejandro D Domínguez-García. Robust distributed average consensus via exchange of running sums. *IEEE Transactions on Automatic Control*, 61(6):1492–1507, 2015.

[3] Ali Jadbabaie, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi. Non-Bayesian social learning. *Games and Economic Behavior*, 76(1):210 – 225, 2012.

[4] Anusha Lalitha, Tara Javidi, and Anand D Sarwate. Social learning and distributed hypothesis testing. *IEEE Transactions on Information Theory*, 64(9):6161–6179, 2018.

[5] Pooya Molavi, Alireza Tahbaz-Salehi, and Ali Jadbabaie. A theory of non-Bayesian social learning. *Econometrica*, 86(2):445–490, 2018.

[6] Angelia Nedić, Alex Olshevsky, and César A Uribe. A tutorial on distributed (non-Bayesian) learning: Problem, algorithms and results. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 6795–6801. IEEE, 2016.

[7] Shahin Shahrampour and Ali Jadbabaie. Exponentially fast parameter estimation in networks using distributed dual averaging. In *52nd IEEE Conference on Decision and Control*, pages 6196–6201. IEEE, 2013.

[8] Artin Spiridonoff, Alex Olshevsky, and Ioannis Ch. Paschalidis. Robust asynchronous stochastic gradient-push: Asymptotically optimal and network-independent performance for strongly convex functions. *Journal of Machine Learning Research*, 21(58):1–47, 2020.

[9] Ye Tian, Ying Sun, and Gesualdo Scutari. Asy-sonata: Achieving linear convergence in distributed asynchronous multiagent optimization. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 543–551. IEEE, 2018.

[10] Tianyu Wu, Kun Yuan, Qing Ling, Wotao Yin, and Ali H Sayed. Decentralized consensus optimization with asynchrony and delays. *IEEE Transactions on Signal and Information Processing over Networks*, 4(2):293–307, 2018.