Discussion of the paper 'A review of distributed statistical inference'

Heng Lian

Taylor & Francis
Taylor & Francis Group

SHORT COMMUNICATION

🔓 OPEN ACCESS Ⓜ Check for updates

Discussion of the paper 'A review of distributed statistical inference'

Heng Lian 🆔

Department of Mathematics, City University of Hong Kong, Hong Kong, People's Republic of China

The authors should be congratulated on their timely contribution to this emerging field with a comprehensive review, which will certainly attract more researchers into this area. In the simplest one-shot approach, the entire dataset is distributed on multiple machines, and each machine computes a local estimate based on local data only, and a central machine performs an aggregation calculation as a final processing step. In more complicated settings, multiple communications are carried out, typically passing also first-order information (gradient) and/or second-order information (Hession matrix) between local machines and the central machine. This review clearly separates the existing works in this area into several sections, considering parameter regression, nonparametric regression, and other models including principal component analysis and variable screening.

In this discussion, I will consider some possible future directions that can be entertained in this area, based on my own personal experience. The first problem is a combination of divide-and-conquer estimation with some efficient local algorithm not used in traditional statistical analysis. This is motivated by that, due to the stringent constraint on the number of machines that can be used either practically or in theory (for example, when using a one-shot approach, the number of machines that can be used is $O(\sqrt{N})$), the sample size on each worker machine can still be large. In other words, even after partitioning, the local sample size may still be too large to be processed by traditional algorithms. In such a case, a more efficient algorithm (one that possibly approximates the exact solution) should be used on each local machine. The important question here is whether the optimal statistical properties can be retained using such an algorithm. One such attempt with an affirmative answer is recently reported in Lian et al. (2021). In this work, we use random sketches (random projection) for kernel regression in an RKHS framework for nonparametric regression. Use of random sketches reduces the computational complexity on each worker machine, and at the same time still retains the optimal statistical convergence rate.

We expect combinations along such a direction can be useful in various settings, and for different settings different efficient algorithms to compute some approximate solution are called for.

The second problem is to extend the studies beyond the worker-server model. Most of the existing methods in the statistics literature are focused on the centralized system where there is a single special machine that communicates with all others and coordinates computation and communication. However, in many modern applications, such systems are rare and unreliable since the failure of the central machine would be disastrous. Consideration of statistical inference in a decentralized system, synchronous or asynchronous, where there is no such specialized central machine, would be an interesting direction of research for statisticians. Currently, decentralized systems are investigated from a purely optimizational point of view, without incorporating statistical properties (Ram et al., 2010; Yuan et al., 2016).

Finally, on the theoretical side, the distributed statistical inference problem provides opportunities and challenges for investigating the fundamental limit (i.e., lower bounds) in performances achievable taking into account communicational, computational and statistical trade-offs. For example, in various models, if a one-short approach is used, then there is a limit in the number of machines allowed in the system and more machines will lead to a suboptimal statistical convergence rate. On the other hand, when multiple communications are allowed, the constraint on the number of machines can be relaxed or even removed. This represents a communicational and statistical trade-off. As another example, the computational and statistical trade-off has already been explored in many works (Khetan & Oh, 2018; L. Wang et al., 2019; T. Wang et al., 2016). The question is how would this change when communications come into play. A general framework taking into account computational, statistical, and communication costs is called for, which would significantly advance the understanding of distributed estimation and inference.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Heng Lian ⬤ http://orcid.org/0000-0002-6008-6614

References

Khetan, A., & Oh, S. (2018). Generalized rank-breaking: Computational and statistical tradeoffs. *Journal of Machine Learning Research*, *19*, 1–42. https://jmlr.org/papers/volume19/16-412/16-412.pdf

Lian, H., Liu, J., & Fan, Z. (2021). Distributed learning for sketched kernel regression. *Neural Networks*, *143*, 368–376. https://doi.org/10.1016/j.neunet.2021.06.020

Ram, S. S., Nedić, A., & Veeravalli, V. V. (2010). Asynchronous gossip algorithm for stochastic optimization: Constant stepsize analysis. In *Recent Advances in Optimization and Its Applications in Engineering* (pp. 51–60). Springer.

Wang, L., Yang, Z., & Wang, Z. (2019). Statistical-computational tradeoffs in high-dimensional single index models. In *Advances in Neural Information Processing Systems* (pp. 10419–10426). The MIT Press.

Wang, T., Berthet, Q., & Samworth, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Annals of Statistics*, *44*(5), 1896–1930. https://doi.org/10.1214/15-AOS1369

Yuan, K., Ling, Q., & Yin, W. (2016). On the convergence of decentralized gradient descent. *SIAM Journal on Optimization*, *26*(3), 1835–1854. https://doi.org/10.1137/130943170