Waheed U. Bajwa, Volkan Cevher, Dimitris Papailiopoulos,
and Anna Scaglione

Machine Learning From Distributed, Streaming Data

The field of machine learning has undergone radical transformations during the last decade. These transformations, which have been fueled by our ability to collect and generate tremendous volumes of training data and leverage massive amounts of low-cost computing power, have led to an explosion in research activity in the field by academic and industrial researchers. Unlike many other disciplines, advances in machine learning research are being rapidly adopted by industry and are beginning to disrupt fields ranging from health care [1], journalism [2], and the retail industry [3] to wireless communications [4], supply-chain management [5], and the automotive industry [6].

In many of the up-and-coming applications of machine learning in these and other fields, such as connected and/or autonomous vehicles, smart grids, edge-caching wireless networks, cloud computing, and urban policing, data are increasingly distributed and are also often streaming. Training predictive models in this distributed, streaming setting requires a rethinking of off-the-shelf machine learning solutions. A number of academic and industrial researchers have recognized the need for this in the last few years; the resulting solutions leverage algorithmic and analytical tools from a number of research areas that cut across multiple disciplines [7]–[10].

Digital Object Identifier 10.1109/MSP.2020.2972654
Date of current version: 28 April 2020

Many of these tools, such as stochastic approximation [11], [12], online learning [13], [14], distributed optimization [15], [16], and decentralized computing [17], have been the mainstay of signal processing researchers for more than a few decades. *IEEE Signal Processing Magazine* (*SPM*), therefore, is one of the best forums for archiving the latest advances in machine learning from data that are distributed, streaming, or both distributed and streaming and for discussing many of the open challenges that remain to be solved for the broad adoption of machine learning tools across a large number of industries that are expected to routinely deal with large volumes of distributed and/or streaming data sets.

Training predictive models in this distributed, streaming setting requires a rethinking of off-the-shelf machine learning solutions.

An overview of the special issue

This special issue of *SPM* on distributed, streaming machine learning presents recent advances in several topic areas that pertain to the training of machine learning models from data that are distributed, streaming, or both distributed and streaming. A particular emphasis of the articles in the special issue is to provide readers with an entry point into algorithmic and analytical techniques that may be relevant to the industry for the emerging era of real-time, decentralized, and autonomous decision making. In particular, the 13 articles not only focus on potentially disruptive techniques that may form the core of future machine learning driven systems, but they also cover techniques that are already being adopted by practitioners. These articles, authored by leading researchers in industry and academia, can be broadly categorized into seven interconnected themes within distributed, streaming machine learning, with some of the articles spanning multiple themes. These themes and their connections to the different overview articles appearing in the special issue are summarized as follows.

Distributed learning

While future machine learning systems will revolve around a number of technological themes, there is one paradigm that is expected to form the core of many future systems. This paradigm, referred to as *distributed learning*, corresponds to an interconnected network of devices/nodes/sites in which each entity has its own set of training data, and the goal is to train a global model that is as accurate as if it had been trained on a single machine that has access to the entire collection of data samples. This paradigm, which is already being extensively explored by academic and industrial researchers, typically arises in applications either where sharing of raw data between different entities cannot take place due to communications or privacy constraints or where the learning task necessarily must be broken across multiple entities due to computational, memory, and/or storage constraints. The articles by Nassif et al., Chang et al., and Cui et al. introduce readers to various aspects of machine learning, which range from general convex and nonconvex learning to the training of specific large-scale models for automatic speech recognition, under this distributed-learning paradigm.

Federated learning

The federated-learning paradigm is somewhat similar to the distributed-learning paradigm in that the data are still distributed across different entities. Unlike the distributed-learning paradigm, however, these different entities (e.g., cell phones, wearable devices, and so on) do not communicate among themselves due to trust issues and/or communications challenges and do not transfer raw data to the cloud due to privacy concerns. Instead, in federated learning, each entity locally updates the global model using its local data and then shares the updated model with a centralized entity, which intermittently passes that model to other entities for further updates and refinements of the global model. The federated-learning paradigm is increasingly gaining popularity, especially within web 2.0 companies, due to privacy reasons. The article by Li et al. provides an overview of the unique characteristics and challenges associated with federated-learning systems.

Learning from streaming data

Streaming is another aspect of modern data sets that will occupy a central place in future machine learning systems. Indeed, many future applications of machine learning are expected to involve data sources that continuously generate data, either at a constant or at a variable rate. Streaming in conjunction with distributed data sets create additional unique challenges that require redesign of many machine learning algorithms. The articles by Koppel et al., Dall'Anese et al., and Xu and Zhao discuss myriad challenges and the corresponding solutions associated with learning from (distributed) data streams under scenarios that range from nonparametric learning and learning in dynamic environments to distributed learning in repeated unknown games.

Distributed optimization for machine learning

Since optimization methods form the bedrock of most machine learning algorithms, distributed optimization is expected to play a major role in machine learning systems that involve distributed data sets. It is in this context that three articles are devoted to a survey of various aspects of distributed optimization that have implications for future machine learning systems. In particular, the article by Nedić provides an overview of distributed-gradient methods for convex learning problems, the article by Xin et al. discusses stochastic first-order methods for distributed machine learning, and the article by Pu et al. explores the role of network topology in distributed stochastic optimization for machine learning.

Distributed reinforcement learning

Another subdomain of machine learning systems that will increasingly have to deal with streaming, distributed data sets is reinforcement learning. Roughly, the basic problem in reinforcement learning, which has a strong overlap with control theory, is to take "actions" based on observed data that maximize some notion of a "reward." Unlike control theory, however, system dynamics are assumed to be unknown in reinforcement learning, and the actor/agent must rely solely on observations for implicit "learning" of the dynamics. Distributed reinforcement learning, in which the streaming observations are also distributed, is almost certain to take center stage in so-called multiagent systems that are abstractions of applications, such as autonomous vehicular networks, autonomous robot swarms, and so on. The article by Lee et al. provides readers with an overview of this emerging area of distributed reinforcement learning.

Coding theory for computations in distributed machine learning

Practical implementations of large-scale distributed machine learning frameworks capable of handling massive data sets also require advances in coding theory for robustness against read/write (storage) errors, computation errors, component failures, communication bottlenecks, and so on. In much the same way as coding-theory techniques enabled the operation of communication systems closer to information theoretic limits, it is expected that a new generation of codes designed for distributed machine learning will enable operation of distributed processing systems closer to their theoretical limits. It is in this vein that the article by Ramamoorthy et al. acquaints the reader with the use of coding theory to mitigate the effects of *stragglers*, defined as slow or failed worker nodes in the system, in distributed matrix computations.

Distributed adversarial machine learning

Given that machine learning systems are expected to be used in critical applications (e.g., management of a nation's power infrastructure and fleets of autonomous vehicles), their robustness and security against adversarial actions and malicious actors become paramount. While the initial focus in this direction has mostly been on centralized problems, recent works have started to develop and analyze algorithms for distributed machine learning systems that can deal with unreliable data, malicious actors, and cyberattacks on individual entities in the network. The article by Yang et al. surveys recent developments pertaining to distributed adversarial machine learning under the threat model of "Byzantine attacks."

Guest Editors

Waheed U. Bajwa (waheed.bajwa@rutgers.edu) received his B.E. (with honors) degree in electrical engineering from the National University of Sciences and Technology,

Pakistan, in 2001 and his M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin–Madison in 2005 and 2009, respectively. He has been with Rutgers University since 2011, where he is currently an associate professor in the Department of Electrical and Computer Engineering and an associate member of the graduate faculty of the Department of Statistics. He received the Army Research Office Young Investigator Award in 2014, the National Science Foundation CAREER Award in 2015, Rutgers University's Presidential Fellowship for Teaching Excellence in 2017, and he was a co-investigator on a work that received the Cancer Institute of New Jersey's Gallo Award for Scientific Excellence in 2017. He is a Senior Member of the IEEE.

Volkan Cevher (volkan.cevher@epfl.ch) received his B.Sc. degree in electrical engineering from Bilkent University, Ankara, Turkey, in 1999 and his Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, in 2005. He is an associate professor at the Swiss Federal Institute of Technology Lausanne and a faculty fellow in the Electrical and Computer Engineering Department at Rice University, Houston. He is a European Laboratory for Learning and Intelligent Systems fellow and was the recipient of the Google Faculty Research Award on Machine Learning in 2018, IEEE Signal Processing Society Best Paper Award in 2016, Best Paper Award at the 2015 IEEE International Workshop of Computational Advances in Multisensor Adaptive Processing, Best Paper Award at the 2009 Signal Processing With Adaptive Sparse Structured Representations Workshop, European Research Council (ERC) Consolidator Grant in 2016, and ERC Starting Grant in 2011. He is a Senior Member of the IEEE.

Dimitris Papailiopoulos (dimitris@papail.io) received his diploma in 2007 and his M.Sc. degree in 2009 from the Technical University of Crete, Greece, and his Ph.D. degree from The University of Texas at Austin in 2014, all in electrical and computer engineering. He is an assistant professor of electrical and computer engineering and computer sciences (by courtesy) at the University of Wisconsin–Madison, faculty fellow of the Grainger Institute for Engineering, and faculty affiliate at the Wisconsin Institute for Discovery. He received the 2019 National Science Foundation CAREER Award, the 2019 Sony Faculty Innovation Award, the 2019 Benjamin Smith Reynolds Award for Excellence in Teaching, and the 2015 IEEE Signal Processing Society Young Author Best Paper Award.

Anna Scaglione (anna.scaglione@asu.edu) received her M.Sc. degree in 1995 and her Ph.D. degree in 1999. She is currently a professor in electrical and computer engineering at Arizona State University. Her research is rooted in statistical signal processing and spans many disciplines that relate to network science, including communication, control, and energy-delivery systems. Her most recent work in signal processing focused on distributed learning and data analytics for signals that are driven by network processes. She is deputy editor-in-chief of *IEEE Transactions on Control Over Networked Systems*. She received the 2000 IEEE Signal Processing Transactions Best Paper Award and the 2013 IEEE Donald G. Fink Prize Paper Award. Her work with her students earned several conference paper awards in addition to the 2013 IEEE Signal Processing Society Young Author Best Paper Award (Lin Li). She is a Distinguished Lecturer for the IEEE Signal Processing Society in 2019–2020. She is a Fellow of the IEEE.

References

[1] S. Gubar, "Using A.I. to transform breast cancer care," *NY Times*, Oct. 24, 2019. [Online]. Available: https://nyti.ms/360AVfr

[2] S. Chandler, "Reuters uses AI to prototype first ever automated video reports," *Forbes*, Feb. 7, 2020. [Online]. Available: http://bit.ly/2v92F3V

[3] J. Greig, "AI, machine learning, robots, and marketing tech coming to a store near you," *TechRepublic*, Feb. 4, 2020. [Online]. Available: https://tek.io/2SgZDSN

[4] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 2019. doi: 10.1109/COMST.2019.2904897.

[5] J. McKendrick, "Artificial intelligence works its way into supply chains," *ZDNet*, Jan. 24, 2020. [Online]. Available: https://zd.net/38zft1D

[6] D. Lu, "AI could help make fast-charging, long-lasting electric car batteries," *New Scientist*, Feb. 19, 2020. [Online]. Available: http://bit.ly/2vb4cXc

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011, doi: 10.1561/2200000016.

[8] B. Recht, C. Re, S. Wright, and F. Niu, "Hogwild!: A lock-free approach to parallelizing stochastic gradient descent," in *Proc. Conf. Neural Information Processing Systems (NeurIPS'11)*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, Eds. 2011, pp. 693–701.

[9] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Consensus-based distributed optimization: Practical issues and applications in large-scale machine learning," in *Proc. 50th Annu. Allerton Conf. Communication, Control, and Computing*, Oct. 2012, pp. 1543–1550. doi: 10.1109/Allerton.2012.6483403.

[10] M. Li et al., "Scaling distributed machine learning with the parameter server," in *Proc. 11th USENIX Symp. Operating Systems Design and Implementation (OSDI'14)*, 2014, pp. 583–598. doi: 10.5555/2685048.2685095.

[11] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, 1951. doi: 10.1214/aoms/1177729586.

[12] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer-Verlag, 2003.

[13] L. Bottou, *Online Learning and Stochastic Approximations*. New York: Cambridge Univ. Press, 2009.

[14] S. Shalev-Shwartz, "Online learning and online convex optimization," *Found. Trends Mach. Learn.*, vol. 4, no. 2, pp. 107–194, 2012. doi: 10.1561/2200000018.

[15] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Automat. Control*, vol. 31, no. 9, pp. 803–812, Sept. 1986. doi: 10.1109/TAC.1986.1104412.

[16] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Automat. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009. doi: 10.1109/TAC.2008.2009515.

[17] R. Bekkerman, M. Bilenko, and J. Langford, Eds., *Scaling Up Machine Learning: Parallel and Distributed Approaches*. New York: Cambridge Univ. Press, 2012.

SP