Hogwild: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent

We're extremely grateful to the NeurIPS 2020 committee for recognizing our paper with the Test of Time award! This project was one we really enjoyed working on together, and we're happy to be here to recollect on this fun project that bridged optimization, computer systems, and machine learning.

In machine learning, we train computers to learn rules or take desired actions by showing them data in the form of many examples, for example labelled pictures of objects and animals. The computer looks at the examples again and again, making small adjustments to its model each time, until eventually it squeezes out as much knowledge as it can gain from the data.

The algorithm used most widely to adjust the model is the stochastic gradient method, known widely as "SGD." Learning from models can use enormous amounts of computation, and it was realized before 2011 that we had to find a way to utilize the power of parallel computing, in particular, we needed to run SGD on parallel computers. But it wasn't obvious how to do this, at least not in a way that yielded good performance and retained the theoretical performance guarantees of SGD.

At the University of Wisconsin-Madison, in early 2011, Chris Re' and Feng Niu had been implementing SGD in parallel, in a fairly conventional way. Each time one of the computer processors looked at some examples and wanted to update the model, it locked access to the model, so that its updates would not interfere with other processors updating at the same time. One day, Feng removed the locks, and found that the resulting "asynchronous" version ran much faster - even though there was the potential for different processors to be updating the model at once. In general, you don't want your parallel computers overwriting each other's work! Bad things usually happen. But Chris and Feng got the same answers as when the processors were forced to carefully take turns. Chris called the resulting approach "Hogwild!", which became the title of our paper. (The exclamation point is significant!)

Chris asked Ben Recht if there was a way to understand this behavior theoretically. As it happened, Ben had talked to me the previous year after I did some analysis of a parallel SGD scheme, based on an important paper of Nemirovski et al that had appeared in 2009. Ben made a crucial addition to the analysis - an assumption that each model update affected only a small part of the model - and suddenly we had theoretical results showing that the asynchronous scheme could (even in theory) achieve a speedup that is linear in the number of parallel processors, to a constant factor - even for quite a large number of processors.

Our NeurIPS 2011 paper contains both analysis and computational results, both of which verify that the Hogwild! approach is effective. Nowadays, asynchronous parallel approaches

are ubiquitous, and our analysis has subsequently been extended in many ways, by us and many others.

We're proud of this paper because the project was such fun to work on, we enjoyed hugely attending the 2011 NeurIPS meeting in Spain, and we were glad to see the work become influential. The paper showed that a "computer systems" perspective on machine learning could produce fruitful results - as has been demonstrated in many later projects. The paper also showcased an "optimization" perspective on machine learning, and this perspective too has proved more and more valuable in the years since 2011. Going forward, we believe that the systems and optimization perspectives will continue to have a lot of value for machine learning. Conversely, we note that machine learning has contributed - and will continue to contribute - to developments in computer systems and optimization.

Author Information

Feng Niu

Feng Niu is a serial entrepreneur who applies machine learning to real world problems. He obtained his PhD from University of Wisconsin-Madison in 2012, and was fortunately there during a time when the three professors who are his coauthors on the Hogwild! Paper were all there. He has co-founded Alation Inc. (data catalog for enterprises) and Lattice Data (acquired by Apple in 2017).

Prof Benjamin Recht. University of California-Berkeley.

Benjamin Recht is an Associate Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. He was previously an Assistant Professor in the Department of Computer Sciences at the University of Wisconsin-Madison. Ben received his B.S. in Mathematics from the University of Chicago, and received a M.S. and PhD from the MIT Media Laboratory. After completing his doctoral work, he was a postdoctoral fellow in the Center for the Mathematics of Information at Caltech.

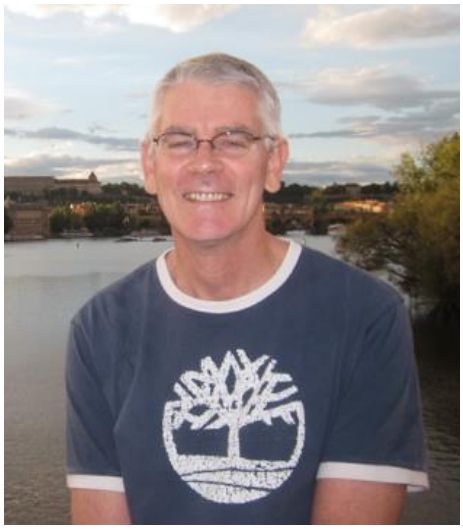
Ben is the recipient of a Presidential Early Career Award for Scientists and Engineers, an Alfred P. Sloan Research Fellowship, the 2012 SIAM/MOS Lagrange Prize in Continuous Optimization, the 2014 Jamon Prize, the 2015 William O. Baker Award for Initiatives in Research, and the 2017 NeurIPS Test of Time Award. He is currently on the Editorial Board of the Journal for Machine Learning Research.

Prof. Christopher Ré. **Stanford University.**

Christopher (Chris) Ré is an associate professor in the Department of Computer Science at Stanford University. He is in the Stanford AI Lab and is affiliated with the Statistical Machine Learning Group. His recent work is to understand how software and hardware systems will change as a result of machine learning along with a continuing, petulant drive to work on math problems. Research from his group has been incorporated into scientific and humanitarian efforts, such as the fight against human trafficking, along with widely used products from technology and enterprise companies including Google Ads, GMail, YouTube, and Apple. He has cofounded four companies based on his research into machine learning systems, SambaNova and Snorkel, along with two companies that are now part of Apple, Lattice (DeepDive) in 2017 and Inductiv (HoloClean) in 2020.

He received a SIGMOD Dissertation Award in 2010, an NSF CAREER Award in 2011, an Alfred P. Sloan Fellowship in 2013, a Moore Data Driven Investigator Award in 2014, the VLDB early Career Award in 2015, the MacArthur Foundation Fellowship in 2015, and an Okawa Research Grant in 2016. His research contributions have spanned database theory, database systems, and machine learning, and his work has won best paper in each area, respectively, at PODS 2012, SIGMOD 2014, and ICML 2016.

Prof. Stephen Wright. University of Wisconsin-Madison.



Stephen Wright holds the George B. Dantzig Professorship, the Sheldon Lubar Chair, and the Amar and Balinder Sohi Professorship of Computer Sciences at the University of Wisconsin-Madison. His research is in computational optimization and its applications to data science and many other areas of science and engineering. Prior to joining UW-Madison in 2001, Wright held positions at North Carolina State University (1986-90), Argonne National Laboratory (1990-2001), and the University of Chicago (2000-2001). He has served as Chair of the Mathematical Optimization Society (2007-2010) and as a Trustee of SIAM (2005-2014). He is a Fellow of SIAM. In 2014, he won the W.R.G. Baker Award from IEEE for best paper in an IEEE archival publication during the three years 2009-2011. He was awarded the Khachiyan Prize by the INFORMS Optimization Society in 2020 for lifetime achievements in optimization.

Prof. Wright is the author / coauthor of widely used text and reference books in optimization including "Primal Dual Interior-Point Methods" and "Numerical Optimization". He has published widely on optimization theory, algorithms, software, and applications.

Prof. Wright served from 2014-2019 as Editor-in-Chief of the SIAM Journal on Optimization and previously served as Editor-in-Chief of Mathematical Programming Series B. He has also served as Associate Editor of Mathematical Programming Series A, SIAM Review, SIAM Journal on Scientific Computing, and several other journals and book series

Media Contact:

Ben Recht
brecht@berkeley.edu