Computer Science > Learning

Divide-and-Conquer Matrix Factorization

Lester Mackey, Ameet Talwalkar, Michael I. Jordan

(Submitted on 5 Jul 2011 (v1), last revised 14 Aug 2012 (this version, v6))

If learning methods are to scale to the massive sizes of modern datasets, it is essential for the field of machine learning to embrace parallel and distributed computing. Inspired by the recent development of matrix factorization methods with rich theory but poor computational complexity and by the relative ease of mapping matrices onto distributed architectures, we introduce a scalable divide-and-conquer framework for noisy matrix factorization. We present a thorough theoretical analysis of this framework in which we characterize the statistical errors introduced by the "divide" step and control their magnitude in the "conquer" step, so that the overall algorithm enjoys high-probability estimation guarantees comparable to those of its base algorithm. We also present experiments in collaborative filtering and video background modeling that demonstrate the near-linear to superlinear speed-ups attainable with this approach.

Comments:	34 pages, 6 figures
Subjects:	**Learning (cs.LG)**; Data Structures and Algorithms (cs.DS); Numerical Analysis (cs.NA); Numerical Analysis (math.NA); Machine Learning (stat.ML)
Cite as:	**arXiv:1107.0789 [cs.LG]**
	(or **arXiv:1107.0789v6 [cs.LG]** for this version)

Submission history

From: Lester Mackey [view email]
[v1] Tue, 5 Jul 2011 06:03:44 GMT (128kb)
[v2] Wed, 17 Aug 2011 00:59:30 GMT (128kb)
[v3] Wed, 21 Sep 2011 01:38:14 GMT (128kb)
[v4] Tue, 1 Nov 2011 05:37:48 GMT (129kb)
[v5] Fri, 18 May 2012 09:28:27 GMT (132kb)
[v6] Tue, 14 Aug 2012 17:33:30 GMT (148kb)

Which authors of this paper are endorsers?