Linear Bandits, Matrix Completion, and Recommendation Systems

Madeleine Udell
Institute for Computational and Mathematical Engineering
Stanford University
Stanford, CA 94305
udell@stanford.edu

Reza Takapoui
Department of Electrical Engineering
Stanford University
Stanford, CA 94305
takapoui@stanford.edu

Abstract

The matrix completion paradigm has received much attention as a solution to the collaborative filtering problem in recommendation systems. The essential idea is that the user-by-item matrix of user preferences may be well modeled as a low rank matrix, and that this assumption may be used to impute unobserved entries, with good performance even when very few entries may be sampled from the matrix [KOM09, CR08]. This literature has focused on minimizing the Frobenius norm error of the recovered matrix from the true matrix, and doing so using a minimal number of random samples from the matrix.

A parallel literature on linearly parametrized bandits has focused on a related problem. Emphasis is placed on the implicit exploration/exploitation tradeoff between exploring the user's true preferences and exploiting the knowledge already acquired to make better decisions [RT10, DM12]. This literature has focused on minimizing the regret incurred by the algorithm using adaptive sampling techniques.

Here, we present first steps towards uniting these two literatures. We give an algorithm using ideas from matrix completion and from the bandit literature to find low regret policies using adaptive sampling when the latent parameters corresponding to both users and items are simultaneously estimated. Recent advances in algorithms for matrix completion [RR11, FNRW11, HKSS12] allow us to efficiently update estimates of the latent parameters of user and items for very large problems. We also give some geometric intuition for our sampling scheme.

We compare two approaches to this problem based on Thompson sampling [Tho33] and upper confidence bounds (UCB) [ACBF02]. While UCB has proved more amenable to theoretical analysis, approaches based on Thompson sampling are often superior in practice, both from a computational standpoint and in an empirical regret comparison [CL12]. We give geometric intuition for the two sampling schemes. Numerical experiments show these methods compare favorably with a naive approach in which a UCB algorithm for the linearly parametrized bandit is applied to each row of the matrix independently.

References

[ACBF02] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[CL12] O. Chapelle and L. Li. An empirical evaluation of thompson sampling. *In ICML Workshop - Online Trading of Exploration and Exploitation 2*, 2012.

[CR08] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *CoRR*, abs/0805.4471, 2008.

[DM12] Yash Deshpande and Andrea Montanari. Linear Bandits in High Dimension and recommendation Systems. *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1750 – 1754, 2012.

[FNRW11] Benjamin Recht Feng Niu, Christopher Ré, and Stephen J. Wright. Hogwild!: A lock-free approach to parallelizing stochastic gradient descent. In *NIPS*, 2011.

[HKSS12] Elad Hazan, Satyen Kale, and Shai Shalev-Shwartz. Near-optimal algorithms for online matrix prediction. *arXiv preprint arXiv:1204.0136*, 2012.

[KOM09] Raghunandan H. Keshavan, Sewoong Oh, and Andrea Montanari. Matrix completion from a few entries. *CoRR*, abs/0901.3150, 2009.

[RR11] Benjamin Recht and Christopher Ré. Parallel stochastic gradient algorithms for large-scale matrix completion. *Optimization Online.*, 2011.

[RT10] P. Rusmevichientong and J. N. Tsitsiklis. Linearly Parameterized Bandits. *Mathematics of Operations Research*, 35(2):395–411, April 2010.

[Tho33] W.R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.