Distributed SGD for matrix factorization

Michal Minařík, Yasin Esfandiari, Yassir Janah
Optimization for Machine Learning

***Abstract*—Low-rank matrix factorization is an effective tool for discovering and analyzing the interactions between two given entities. It has been successfully applied in recommendation system (users and items) or news personalization (users and articles) [1]. In production systems, matrices with millions of rows and columns can be involved [2]. Therefore, efficient algorithms that approximate the solution in a reasonable time need to be implemented, often leveraging distributed computing [3], [2]. In this work, one such algorithm using distributed stochastic gradient descent [1] is implemented in PySpark [4] and analyzed. The analysis covers convergence time of the recommendation system task for different number of GPU nodes, on available datasets like Netflix [5]. Moreover, a general analysis of distributed SGD behavior is presented [6], [7], [8], [9].**

References

[1] R. Gemulla, E. Nijkamp, P. Haas, and Y. Sismanis, "Large-scale matrix factorization with distributed stochastic gradient descent," 08 2011, pp. 69–77. [Online]. Available: https://doi.org/10.1145/2020408.2020426

[2] A. S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: Scalable online collaborative filtering," in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 271–280. [Online]. Available: https://doi.org/10.1145/1242572.1242610

[3] C. Liu, H.-c. Yang, J. Fan, L.-W. He, and Y.-M. Wang, "Distributed nonnegative matrix factorization for web-scale dyadic data analysis on mapreduce," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 681–690. [Online]. Available: https://doi.org/10.1145/1772690.1772760

[4] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin, A. Ghodsi, J. Gonzalez, S. Shenker, and I. Stoica, "Apache Spark: A unified engine for big data processing," *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, nov 2016.

[5] J. Bennett, S. Lanning, and N. Netflix, "The netflix prize," in *In KDD Cup and Workshop in conjunction with KDD*, 2007.

[6] X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, W. Zhang, and J. Liu, "Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/f75526659f31040afeb61cb7133e4e6d-Paper.pdf

[7] E. Yu, D. Dong, Y. Xu, S. Ouyang, and X. Liao, "Cd-sgd: Distributed stochastic gradient descent with compression and delay compensation," in *50th International Conference on Parallel Processing*, ser. ICPP 2021. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3472456.3472508

[8] S. Ahn, A. K. Balan, N. Liu, S. Rajan, and M. Welling, "Large-scale distributed bayesian matrix factorization using stochastic gradient MCMC," *CoRR*, vol. abs/1503.01596, 2015. [Online]. Available: http://arxiv.org/abs/1503.01596

[9] B. Swenson, R. Murray, H. V. Poor, and S. Kar, "Distributed stochastic gradient descent: Nonconvexity, nonsmoothness, and convergence to local minima," 2020. [Online]. Available: https://arxiv.org/abs/2003.02818